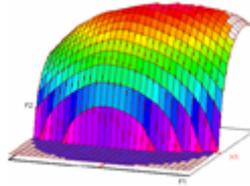


Année Universitaire : 2017-2018



**Master Sciences et Techniques CAC Ageq  
Chimométrie et Analyse Chimique : Application à la gestion de la qualité**

**MEMOIRE DE FIN D'ETUDES  
Pour l'Obtention du Diplôme de Master Sciences et Techniques**

**Application de la régression logistique à l'évaluation de la  
qualité chimique des eaux d'irrigation dans la région de  
Skhirat**

**Présenté par:**

**HAMID RAJAA**

**Encadré par:**

- **Dr. Ahmed DOUIK (INRA-Rabat)**
- **Pr. Fouad OUAZZANI (FST-Fès)**

**Soutenu Le 13 Juin 2018 devant le jury composé de:**

- **Dr. Ahmed DOUIK**
- **Pr. Fouad OUAZZANI**
- **Pr. Abdellah FARAH**
- **Pr. Khadija MOUGHAMIR**

**Stage effectué à : INRA – Rabat**

## *DÉDICACES*

*Je dédie ce travail*

*A mes parents*

*A mon mari*

*En témoignage de ma gratitude pour tant*

*De sacrifices, de prières*

*D'amour et d'encouragement.*

*Que dieu leur accorde une longue vie.*

*A mon frère et mes sœurs*

*Pour leurs encouragements et leurs*

*affections.*

*A mes amies et collègues*

*Pour tous les bons moments qu'on a passé*

*Ensemble.*

*Avec mes souhaits d'un avenir plein de*

*Joie et de succès.*

## ***Remerciements***

Je tiens à remercier le chef du Centre Régional de la Recherche Agronomique (CRRA) de Rabat, **Mr. Abdellah El HOUSNI** qui m'a bien accueilli au sein de son centre, en m'accordant un stage à l'Unité de Recherche sur l'Environnement et la Conservation des Ressources Naturelles (URECRN). Je le remercie aussi pour sa gentillesse et son soutien.

Mes profonds remerciements vont à mon encadrant de stage, **Mr. Ahmed DOUAIK**, Docteur en Agro-climatologie, Géomatique, Salinité du sol, Statistique et Géostatistique à l'URECRN au CRRA de Rabat. Il a conduit ce stage d'une haute compétence et grande disponibilité. J'ai particulièrement été impressionné par ses qualités scientifiques et humaines ainsi que par sa patience, sa gentillesse et la rigueur avec laquelle il a mené à bien ce travail.

Mes remerciements profonds s'étendent également à mon encadrant, **Mr. Fouad OUAZZANI**, professeur à la Faculté de Sciences et Techniques de Fès pour sa gentillesse, sa disponibilité incessante, son aide et ses conseils précieux.

Je tiens à remercier aussi les membres du jury : **Mr. Abdellah FARAH** et **Mme. Khadija MOUGHAMIR** qui m'ont fait l'honneur d'examiner ce travail et de participer à ce jury.

Hommages respectueux et sincères à **Mr El Mestafa EL HADRAMI** : Responsable du Master CAC Agiq pour les efforts qu'il a consenti en faveur de ma formation, et pour son appréciable aide, ainsi que pour la qualité de son enseignement.

J'exprime également ma gratitude à l'ensemble des enseignants de la formation CAC agiq pour avoir porté un vif intérêt à ma formation, et pour avoir accordé le maximum de leur temps, leur attention et leur énergie, dans un cadre très agréable de complicité et de respect.

Je tiens à témoigner ma reconnaissance et ma gratitude à **Mr. Abdelmjid ZOUAHRI**, Coordinateur de l'URECRN au CRRA de Rabat.

Mes remerciements sont également adressés aux techniciens des laboratoires de l'URECRN ainsi qu'aux personnels du CRRA de Rabat pour leur gentillesse et leur disponibilité à rendre service.

# Sommaire

<u>Introduction générale</u> .....	1
 <u>PARTIE 1 : Bibliographie</u>	
CHAPITRE I : Présentation de lieu de stage	
I-    INRA .....	2
II-   Centre Régional de la Recherche Agronomique (CRRA) de Rabat .....	2
III-  Unité de Recherche sur l'Environnement et la Conservation des Ressources Naturelles (URECRN).....	4
CHAPITRE II : Généralités sur l'irrigation	
I-    Introduction.....	6
II-   Techniques d'irrigation au service du développement agricole .....	6
III-  Impacts de la qualité chimique des eaux d'irrigation sur le sol et les cultures .....	9
CHAPITRE III: Régression logistique	
I-    Définition de la régression logistique.....	13
III-  Intérêts de la régression logistique .....	14
IV-   Définition mathématique du modèle de régression logistique .....	15
V-    Domaines d'application de la régression logistique.....	23
 <u>PARTIE 2: Matériel et méthodes</u>	
CHAPITRE I : Etude expérimentale	
I-    Description de la zone d'étude .....	24
II-   Méthodes d'échantillonnage de l'eau .....	24
III  Analyse des paramètres physico-chimiques .....	25
IV  Normes de qualité des eaux destinées à l'irrigation .....	28
CHAPITRE II : Logiciel SPSS	
I-    Présentation du logiciel SPSS .....	29
II-   Méthodes de régression logistique dans SPSS .....	29
 <u>PARTIE 3 : Résultats et discussion</u>	
CHAPITRE I : Statistique descriptive .....	31
CHAPITRE II : Régression logistique .....	38
<u>Conclusions</u> .....	46
<u>Références bibliographiques</u> .....	47
<u>Annexes</u> .....	48
<u>Résumé</u>	

## *Liste des tableaux*

Tableau 1. Tableau comparatif des efficacités des techniques d'irrigation .....	8
Tableau 2. Normes indicatives de la qualité de l'eau d'irrigation .....	28
Tableau 3. Paramètres statistiques descriptifs de l'eau d'irrigation .....	31
Tableau 4. Coefficients de corrélation de Pearson entre les différents paramètres chimiques .....	32
Tableau 5. Tableau des tests de spécification du modèle de régression SAR binaire .....	38
Tableau 6. Tableau de test de Hosmer-Lemeshow pour modèle de régression SAR binaire .....	38
Tableau 7. Tableau des variables dans l'équation pour modèle de régression SAR binaire.....	39
Tableau 8. Tableau de classement pour modèle de régression SAR <sub>binaire</sub> .....	39
Tableau 9. Tableau des informations d'ajustement de modèle de régression SAR ordinal .....	40
Tableau 10. Tableau des estimations de paramètre du modèle régression SAR <sub>ordinal</sub> .....	41
Tableau 11. Tableau croisé SAR ordinal * Modalité de réponse prévue .....	42
Tableau 12. Tableau des tests de spécification du modèle de base de CE binaire.....	42
Tableau 13. Tableau de test de Hosmer-Lemeshow pour le modèle de base de CE <sub>binaire</sub> .....	43
Tableau 14. Tableau des variables dans l'équation de modèle de base de CE binaire .....	43
Tableau 15. Tableau de classement pour le modèle de régression de CE binaire.....	43
Tableau 16. Tableau des informations d'ajustement de modèle de régression de CE <sub>ordinal</sub> .....	44
Tableau 17. Tableau des estimations des paramètres pour le modèle de régression de CE <sub>ordinal</sub> ...	44
Tableau 18. Tableau croisé CE ordina * prédiction CE <sub>ordinal</sub> .....	45

## *Liste des figures*

Figure 1 : Différents systèmes d'irrigation.....	7
Figure 2 : Exemple de modèle de régression logistique .....	16
Figure 3. Diagramme de dispersion SAR-Ca <sup>2+</sup> .....	33
Figure 4. Diagramme de dispersion SAR-Mg <sup>2+</sup> .....	33
Figure 5. Diagramme de dispersion SAR-Na <sup>+</sup> .....	34
Figure 6. Diagramme de dispersion SAR-CO <sub>3</sub> <sup>--</sup> .....	34
Figure 7. Diagramme de dispersion CE-pH.....	35
Figure 8. Diagramme de dispersion CE-Ca <sup>2+</sup> .....	35
Figure 9. Diagramme de dispersion CE-Mg <sup>2+</sup> .....	36
Figure 10. Diagramme de dispersion CE-Na <sup>+</sup> .....	36
Figure 11. Diagramme de dispersion CE-K <sup>+</sup> .....	36
Figure 12. Diagramme de dispersion CE-Cl <sup>-</sup> .....	37
Figure 13. Diagramme de dispersion CE-SO <sub>4</sub> <sup>--</sup> .....	37

## *Liste des abréviations*

CE : Conductivité électrique  
CRRA : Centre régional de la recherche agronomique  
Ddl: Degré de liberté  
INRA : Institut national de la recherche agronomique  
GLM: Generalized Linear Model  
L: likelihood  
LL: log-likelihood  
LR: likelihood-ratio  
OR: odds ratio  
PASW: Predictive Analytics Software  
pH : Potentiel hydrogène  
RAS : Rapport d'absorption de sodium  
RCB : Risque lié aux carbonates et aux bicarbonates  
SAR : Sodium absorption ratio  
S.E: Standard error  
SIG : Système d'information géographique  
SIGVA : Système d'Information Géographique de Vocation Agricole  
STEP : STation d'Epuration des eaux usées  
SST: Sels solubles totaux  
SPSS: Statistical Package for the Social Sciences  
URECRN : Unité de Recherche sur l'Environnement et la Conservation des Ressources

# *Introduction générale*

Partout dans le monde, la pression sur les ressources en eau, et en particulier sur les ressources en eau souterraines, est à la hausse, principalement en raison de la demande croissante et de la dégradation de la qualité de l'eau. L'accès généralisé à l'eau potable, l'irrigation, l'expansion urbaine, le développement industriel sont autant des facteurs qui font augmenter ces pressions.

Au Maroc, la disponibilité en eaux est, jusqu'à présent, très limitée et risquerait de diminuer fortement à long terme en raison des périodes de sécheresse prolongées, la croissance démographique et le changement climatique qui font plus pression sur cette ressource. En outre, le pompage de l'eau de zones côtières et le niveau de la mer en hausse en raison du changement climatique peuvent détériorer la qualité de l'eau.

La qualité de l'eau d'irrigation est un paramètre essentiel pour le rendement des cultures, le maintien de la productivité du sol et la protection de l'environnement. Elle peut être mieux déterminée par une analyse chimique en laboratoire. On met en évidence quatre critères d'évaluation de la qualité de l'eau pour irrigation qui sont : la teneur totale en sels solubles (risque de salinité), la proportion relative d'ions de sodium par rapport aux autres cations (risque de sodium - effets sur la perméabilité du sol), alcalinité liée à la concentration de calcium et de magnésium (alcalinité) et la concentration des éléments toxiques (risque de toxicité).

Au Maroc, les conditions naturelles ont rendu l'irrigation une exigence technique clé qui a acquis d'indéniables dimensions économiques et sociales.

La région de Skhirat possède d'énormes réserves d'eau souterraine. Elle est caractérisée par une importante agriculture intensive. En conséquence les eaux souterraines seraient de mauvaise qualité et de quantité. Puisque la surexploitation de ces ressources conduit à la dégradation de la qualité de l'eau et de sol, ceci entraîne des problèmes de pollution par les nitrates, la salinisation et la solidification des sols.

Donc il est nécessaire de réaliser une étude pour évaluer la qualité de ces eaux d'irrigation. C'est dans ce sens que cette étude est faite pour évaluer les eaux d'irrigation de la région de Skhirat par des analyses physico-chimiques et par l'application de la régression logistique.

# **PARTIE 1 : Bibliographie**

## **CHAPITRE I : Présentation du lieu de stage**

### **I- INRA**

L'institut National de la recherche Agronomique « INRA » a pour **mission** d'entreprendre les recherches pour le développement agricole. C'est un établissement public dont les origines remontent à 1914 avec la création des premiers services de recherche agricole officiel. Il a connu en 2003 une **réorganisation structurelle** visant la modernisation de son processus de gestion.

La finalité de la nouvelle organisation est de doter l'institution:

- d'une planification stratégique adéquate pour renforcer les capacités prospectives d'adaptation, de réaction et d'anticipation de la demande sociale de recherche agronomique;
- d'une politique de proximité en se basant sur la régionalisation et la déconcentration de la recherche;
- d'un système intégré de suivi, d'évaluation et de contrôle;
- d'une gestion intégrée et rationnelle des ressources;
- d'une politique de valorisation de ses produits;
- d'une politique cohérente d'information et de coopération.

L'INRA opère à travers dix **centres régionaux** de la recherche agronomique et 23 domaines expérimentaux répartis sur le territoire national et couvrant les divers agrosystèmes du pays.

Les projets de recherche de l'INRA sont définis avec la participation des partenaires, des clients et des prescripteurs régionaux. Ils sont menés au sein de trente unités de recherche hébergés par les Centres Régionaux. Ils sont encadrés à l'échelle centrale par dix départements scientifiques à vocation disciplinaire.

Pour accomplir sa mission et être au diapason de l'actualité scientifique, l'INRA entretient des relations de partenariats avec des organisations nationales et internationales, les structures de développement, le secteur privé et les Organisations Non Gouvernementales [1].

### **II- Centre Régional de la Recherche Agronomique (CRRA) de Rabat**

Le Centre régional de la Recherche Agronomique de Rabat a été créé en 2003. Il est chargé de mener les études scientifiques, techniques et économiques pour le développement de l'agriculture et de l'élevage et la conservation des ressources naturelles au niveau de la zone d'action du Centre.

- **Organisation :**

Le Centre Régional de la Recherche Agronomique de Rabat est organisé en quatre types de structures selon la nature des missions qui leurs sont assignées [2] :

### 1- **Structures de recherche**

- Unité de Recherche sur l'Amélioration des Plantes, Conservation et Valorisation des Ressources Phytogénétiques, regroupant :

- Laboratoire d'amélioration des plantes (Céréales, plantes aromatiques et médicinales, cultures fourragères, betterave à sucre),
- Laboratoire de cytogénétique
- Laboratoire de germoplasme
- Laboratoire de microbiologie
- Laboratoire d'agro-physiologie
- Laboratoire de protection des plantes.

- Unité de Recherche sur l'Environnement et la Conservation des Ressources Naturelles, regroupant :

- Laboratoire d'agro-climatologie
- Laboratoire d'analyses de sol
- Laboratoire de salinité et de nutrition minérale des plantes.

- Unité de recherche en Biotechnologie, regroupant :

- Laboratoire de culture in vitro
- Laboratoire de marqueurs moléculaires et de transformation génétique.

- Unité de Recherche sur la Production Animale et Fourragère, regroupant :

- Laboratoire de nutrition animale
- Laboratoire d'amélioration génétique animale
- Laboratoire de physiologie végétale et agronomie
- Laboratoire de technologie agro-alimentaire et de qualité.

### 2- **Structure de liaison**

- Service de Recherche – Développement

### 3- **Structures d'appui à la recherche**

- Service Administratif
- Domaine Expérimental de Marchouch : 1060 ha
- Domaine Expérimental d'El Koudia : 880 ha
- Domaine Expérimental du Jardin d'Essais Botaniques: 17 ha

### 4- **Structures de concertation**

- Conseil Régional Consultatif d'Orientation de la Recherche
- Comité Scientifique et de Gestion.

### **III- Unité de Recherche sur l'Environnement et la Conservation des Ressources Naturelles (URECRN)**

#### **1- Présentation**

L'Unité de Recherche sur l'Environnement et la Conservation des Ressources Naturelles, affiliée au Centre Régional de la Recherche Agronomique de Rabat, a des missions régionale et nationale de gestion durable des ressources naturelles (sol, eau et climat). Elle est dotée d'infrastructure et d'équipements importants: Serre, laboratoires d'analyse, laboratoire de SIG et cartographie, station météorologique automatique, ...

Les thèmes de recherche traités au niveau de l'Unité sont :

- Carte de vocation agricole des terres,
- Agro-climatologie ;
- Gestion et conservation du sol et de l'eau (semis direct, salinité, pollution nitrique, ...) ;
- Fertilité du sol et fertilisation des cultures (carte de fertilité des sols et normes de fertilisation des cultures) [1].

#### **2- Activités :**

Les principales activités de recherche sont :

- Réalisation des cartes de vocation agricole des terres pour plusieurs provinces du Maroc (Environ 6 million ha) ;
- Evaluation de l'aptitude des sols pour les cultures des régions étudiées ;
- Elaboration des bases de données sur le sol, le climat et les principales cultures ;
- Elaboration d'un Système d'Information Géographique de Vocation Agricole (SIGVA) pour les régions étudiées ;
- Réalisation de l'Atlas Agro-climatique du Maroc ;
- Détermination de la durée de la période de croissance (LGP) pour les cultures en zones bour ;
- Quantification de la fréquence des événements extrêmes ;
- Evaluation de l'impact du changement climatique sur la longueur de la période de croissance et sur le potentiel de rendement des cultures ;
- Etude des paramètres de qualité des sols (fertilité, salinité, alcalinité, ...) des zones maraîchères de Skhirat, Ben Slimane, Tiflet et Khémisset ;
- Etude de la qualité des eaux d'irrigation (salinité et pollution nitrique) des zones maraîchères de Skhirat, Ben Slimane, Tiflet et Khémisset ;
- Quantification de la lixiviation de l'azote nitrique et des sels dans la zone de Béni Amir (Tadla) et de Skhirat ;

- Caractérisation de la qualité des eaux usées épurées de la STEP de Skhirat et quantification des apports azotés de ces eaux usées en irrigation ;
- Diagnostic du besoin et réponse à l'inoculation de l'haricot vert chez les agriculteurs de la région de Rabat-Salé-Zemmour-Zaer ;
- Effet de la technologie du semis direct sur les paramètres hydrodynamiques et les facteurs d'érosion des sols ;
- Effet de la technologie du semis direct sur les rendements des céréales et légumineuses alimentaires ;
- Evaluation de la fertilité des sols et établissement des cartes de fertilité des sols des zones de Rabat, Khémisset et Sidi Kacem ;
- Etablissement des normes de fertilisation des cultures : blé, lentille et pomme de terre dans la région de Rabat-Salé-Zemmour-Zaer [1].

## **CHAPITRE II : Généralité sur l'irrigation**

### **I- Introduction**

L'agriculture est, de loin, l'industrie ayant la plus grande consommation d'eau. L'irrigation des régions agricoles représente 70% de l'eau utilisée dans le monde entier.

Dans plusieurs pays en voie de développement, l'irrigation représente jusqu'à 95% de toutes les utilisations d'eau, et joue un rôle important dans la production de nourriture et la sécurité alimentaire. Les futures stratégies de développement agricole de la plupart de ces pays dépendent de la possibilité de maintenir, d'améliorer et d'étendre l'agriculture irriguée.

D'autre part, il existe une pression croissante sur les ressources en eau, amplifiée par la concurrence des autres secteurs utilisateurs d'eau et par le respect de l'environnement.

L'eau est une ressource qui peut créer des tensions entre différents pays se partageant les mêmes sources d'eau. L'agriculture irriguée peut entraîner une grande concurrence, puisqu'elle représente de 70 à 90% de l'utilisation d'eau dans certaines régions.

Chacun connaît le rôle primordial que joue l'eau dans la vie des plantes. Elle transporte en effet les éléments minéraux, qui, puisés par les racines dans le sol, vont alimenter l'extraordinaire usine chimique qu'est le végétal.

L'eau d'irrigation est prélevée sur un point d'eau (rivière, lac ou nappe aquifère) et conduite jusqu'au champ grâce à une infrastructure de transport appropriée. Pour satisfaire leurs besoins en eau, les cultures irriguées bénéficient à la fois de l'apport d'eaux de pluie naturelles plus ou moins fiables et de celui d'eaux d'irrigation. L'irrigation constitue un outil de gestion efficace contre les aléas des précipitations. Elle permet de choisir des variétés à haut rendement en appliquant les fertilisants nécessaires, ainsi que les traitements de lutte contre les ravageurs et d'autres intrants, et rend ainsi ces cultures économiquement intéressantes. Elle a pour effet de favoriser l'augmentation des rendements [3].

### **I- Techniques d'irrigation au service du développement agricole**

L'irrigation est une technique artificielle utilisée sur des terres agricoles, dans le but d'augmenter leurs rendements, surtout quand il y a un manque ou une insuffisance d'eau de pluie.

L'irrigation permet d'optimiser les rendements des terres agricoles, par contre elle nécessite l'utilisation de grandes quantités d'eau, à travers l'apport de l'eau nécessaire à une forte croissance.

L'irrigation peut aussi avoir d'autres applications :

- l'apport d'éléments fertilisants soit au sol, soit, par aspersion, aux feuilles (fertilisation foliaire); dans la culture hydroponique, l'irrigation se confond totalement avec la fertilisation ;

- la lutte contre le gel, par aspersion d'eau sur le feuillage (vergers, vignobles) peut permettre de gagner quelques degrés de température précieux au moment des gelées printanières, voire dans certains cas par inondation.

Cette technique vise à couvrir le manque d'eau d'origine naturelle (pluie ou nappe phréatique), surtout dans les zones désertiques ou au climat très sec, dans lesquelles les précipitations sont rares. L'irrigation est également appliquée dans le cas des cultures qui demandent plus d'eau qu'elles ne peuvent en trouver sur place, telles que le maïs, le coton, ...

## II-1 Techniques d'irrigation

L'irrigation est un apport artificiel d'eau aux cultures en complément des précipitations naturelles. Le but de cette technique est de créer des conditions favorables de production. On classe ces systèmes en trois types comme on peut le voir sur la Figure 1.

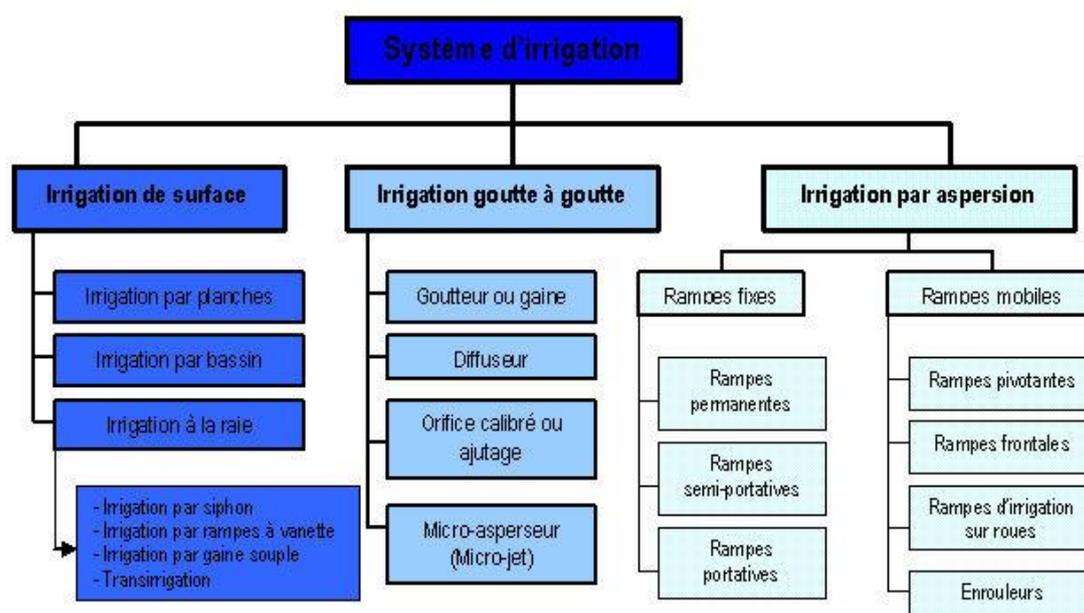


Figure 1 : Différents systèmes d'irrigation.

**A- Irrigation par écoulement de surface :** aussi appelée **irrigation gravitaire**, consiste à distribuer l'eau par le biais de canaux et de rigoles. Sous l'effet de la gravité, les canaux distribuent l'eau à des canaux secondaires jusqu'aux parcelles. Comme tout le dispositif est à ciel ouvert, beaucoup d'eau est perdue par absorption du sol et évaporation.

**B- Irrigation par micro-irrigation (goutte à goutte) :** La micro-irrigation se fait au **goutte-à-goutte**. L'eau est amenée directement au pied de la plante. Le système est très économe en eau puisqu'il ne consomme que ce dont la plante a besoin. Le réglage doit être très bien fait, ce qui n'est pas aisé. L'eau utilisée doit être filtrée afin de ne pas boucher les goutteurs.

**C- Irrigation par aspersion :** Cette irrigation se réalise dans le cadre d'un système d'arrosage intégré, ou enterré. L'eau circule dans des canalisations enterrées. Elle sort vers des tuyaux mobiles qui la distribuent aux cultures via des systèmes d'aspersion. L'irrigation par aspersion imite une pluie fine [4].

## II-2: Efficacité des techniques d'irrigation

L'efficacité de ces techniques est mesurée par le rapport suivant :

$$\text{Efficience} = \frac{\text{Volume d'eau utilisé par les plantes}}{\text{Volume total d'eau fournie}}$$

Actuellement, l'efficience moyenne des systèmes d'irrigation mondiaux est de 40%. Ceci est dû à plusieurs facteurs :

- réseaux de distribution mal entretenus et/ou mal dimensionnés
- maintenance inadéquate
- mauvaise organisation institutionnelle
- manque de formation auprès des agriculteurs [5].

On peut voir dans le Tableau 1 les différences de rendement entre les trois techniques d'irrigation.

Tableau 1. Tableau comparatif des efficacités des techniques d'irrigation [5].

Système d'irrigation	Gravitaire	Par aspersion	Locale
Efficience	20 à 60 %	65 à 85 %	85 à 95%

→ L'irrigation en chiffres

- **17%** des terres cultivées, soit **270 millions d'hectares** (ha), sont irriguées
- **1/3** de la production alimentaire mondiale
- un champ irrigué produit jusqu'à **3,5 fois plus**
- un accroissement de plus de **200%** des surfaces irriguées en **40 ans**
- Inde, Chine et Etats-Unis (respectivement **57, 50 et 21 millions d'ha**) représentent **50%** des surfaces irriguées totales

L'accroissement de la production alimentaire, nécessaire au vu de l'augmentation de la population mondiale, passera principalement par une généralisation et une optimisation des systèmes d'irrigation [5].

## **II- Impacts de la qualité chimique des eaux d'irrigation sur le sol et les cultures**

Au Maroc, la disponibilité en eaux est, jusqu'à présent, très limitée et risquerait de diminuer fortement à long terme en raison des périodes de sécheresse prolongées. En effet, les estimations prévisionnelles effectuées par la banque mondiale en 1994 ont déclenché la sonnette d'alarme. En effet, les ressources hydriques renouvelables par habitant risqueraient de diminuer de moitié en passant de 800 m<sup>3</sup> en 1990 à 400 m<sup>3</sup> en 2020, classant ainsi le Maroc dans la catégorie des pays en situation de stress hydrique chronique. La détérioration de la qualité des ressources en eau par prolifération des différentes sources de pollution (engrais et pesticides, rejets d'eaux usées non traitées, rejets de déchets solides non contrôlés, exploitation minière, urbanisation, etc.) constitue une menace aussi importante que celle liée au déséquilibre quantitatif.

La connaissance de la qualité de l'eau utilisée pour l'irrigation est un élément essentiel en agriculture pour le suivi des cultures en fonction de leurs besoins, mais aussi pour son impact sur l'environnement et le risque de salinisation des sols.

La qualité de l'eau d'irrigation peut être mieux déterminée par une analyse chimique en laboratoire. Selon Ayers et Westcot (1989), les critères d'évaluation de la qualité de l'eau d'irrigation sont les suivants:

- Salinité: Contenu total en sels solubles
- Sodium : Proportion relative des cations sodium (Na<sup>+</sup>) par rapport aux autres
- Alcalinité et dureté: Concentration d'anions carbonates (CO<sub>3</sub><sup>2-</sup>) et bicarbonates (HCO<sub>3</sub><sup>-</sup>) en relation avec la concentration en calcium (Ca<sup>2+</sup>) et en magnésium (Mg<sup>2+</sup>)
- Concentration en éléments qui peuvent être toxiques
- pH de l'eau d'irrigation [6]

### **1- Salinité**

Les principaux sels responsables de la salinité de l'eau sont les sels de calcium (Ca<sup>2+</sup>), de magnésium (Mg<sup>2+</sup>), de sodium (Na<sup>+</sup>), les chlorures (Cl<sup>-</sup>), les sulfates (SO<sub>4</sub><sup>2-</sup>) et les bicarbonates (HCO<sub>3</sub><sup>-</sup>). Une valeur élevée de la salinité signifie une grande quantité d'ions en solution, ce qui rend plus difficile l'absorption de l'eau et des éléments minéraux par la plante.

La salinisation est un processus d'accumulation de sels à la surface du sol et dans la zone racinaire qui occasionne des effets nocifs sur les végétaux et le sol, il s'en suit une diminution des rendements, et à terme, une stérilisation du sol.

Ce phénomène est un terme générique caractérisant une augmentation progressive de la concentration des sels dans la solution du sol conduisant ainsi à la précipitation successive de minéraux qui modifie sa composition et détermine différentes voies d'évolution des sols en fonction de l'abondance relative des différents ions majeurs dans la solution.

## 2- Sodium

Le sodium est l'un des éléments les plus indésirables dans l'eau d'irrigation. Cet élément trouve son origine dans l'altération de la roche et du sol, des intrusions d'eau de mer, des eaux traitées et des systèmes d'irrigation.

Le problème principal avec une grande quantité de sodium est son effet sur la perméabilité du sol et sur l'infiltration de l'eau. Le sodium remplace le calcium et le magnésium adsorbés sur les particules d'argile et provoque la dispersion des particules du sol. Il y a donc éclatement des agrégats du sol ce qui provoque un sol dur et compact lorsqu'il est sec et excessivement imperméable à l'eau. La perméabilité des sols sableux peut ne pas se détériorer aussi vite que les sols plus lourds lorsqu'irrigués avec une eau de forte teneur en sodium, cependant un problème potentiel existe.

Le sodium contribue aussi directement à la salinité totale de l'eau et peut être toxique pour des cultures sensibles comme les carottes, les haricots, les fraises, les framboises, les oignons, pour en nommer quelques-unes.

De grande quantité de sodium combiné avec du chlorure donne à l'eau un goût salé. Si l'eau passe par un système d'aspersion et que le calcium et le magnésium sont bas, des teneurs moyennes à élevées de sodium peuvent défolier les plantes sensibles.

La concentration de sodium dans l'eau d'irrigation est estimée par le ratio d'absorption du sodium (SAR : Sodium absorption ratio). Le SAR décrit la quantité de sodium en excès par rapport aux cations calcium et magnésium, qui eux, peuvent être tolérés en relativement grande quantité dans l'eau d'irrigation :

$$SAR = \frac{Na^+}{\sqrt{(Ca^{2+} + Mg^{2+})/2}}$$

On utilise le SAR pour classifier les risques liés au sodium dans les sources d'eau d'irrigation. L'eau caractérisée par un SAR supérieur à 10 aura tendance à produire une accumulation de sodium dans le sol. Le sodium agit au niveau de la défloculation du sol argileux ce qui entraîne une diminution de la macro-porosité (air) et du taux d'infiltration de l'eau.

## 3- Alcalinité et dureté

### 3-1 Alcalinité

L'alcalinité est une mesure du pouvoir de l'eau à neutraliser les acides, c'est un peu comme le «pouvoir tampon de l'eau». En d'autres termes, l'alcalinité mesure la résistance à tout changement de pH. Le pouvoir neutralisant de l'eau est attribué principalement à la présence de bicarbonates de calcium et de magnésium dissous dans l'eau (également dans une moindre mesure, des hydroxydes, des bases organiques, des borates, des ions ammoniums, phosphates et silicates).

Tant que l'acide n'a pas maîtrisé les ions responsables de l'alcalinité, le pH ne bouge pratiquement pas.

Une fois le processus d'acidification enclenché, la réaction produit alors de l'eau (H<sub>2</sub>O), du dioxyde de carbone (CO<sub>2</sub>) qui est libéré dans l'air ainsi que le cation accompagnateur (soit le calcium ou le magnésium).

### **3-2 Dureté**

La dureté fait référence à la quantité de calcium et de magnésium contenue dans l'eau. Ces deux éléments proviennent de l'altération de la roche-mère.

La teneur en calcium est habituellement plus élevée que le magnésium dans les eaux souterraines. Par contre, là où on trouve de la contamination par l'eau de mer, la concentration en magnésium peut être plus élevée que celle du calcium.

En général, une eau dure n'est pas souhaitable pour l'usage domestique car elle réduit de beaucoup l'efficacité des savons. Par contre, en agriculture, une eau dure aide le sol à conserver une bonne structure par les ponts calciques.

Lorsque l'alcalinité est élevée, le calcium et le magnésium sont les deux principaux éléments qui causent des problèmes d'eau incrustante et qui peuvent progressivement obstruer les goutteurs par des dépôts calcaires.

Il arrive toutefois que la dureté de l'eau soit élevée sans pour autant que son alcalinité le soit. C'est le cas lorsque le calcium et le magnésium de l'eau sont sous forme de sulfates et de chlorures.

## **4. Eléments toxiques**

Certains éléments de l'eau d'irrigation peuvent être directement toxiques à la culture. Établir des limites de toxicité pour l'eau d'irrigation est compliqué de par les réactions qui peuvent se passer quand l'eau atteint le sol. Les éléments potentiellement dangereux de l'eau peuvent être inactivés par des réactions chimiques ou bien s'accumuler dans le sol jusqu'à atteindre des niveaux de toxicité pour les plantes.

### ***4.1 Chlorures et sulfates***

Lorsqu'ils sont présents dans l'eau d'irrigation, ces éléments contribuent à augmenter la concentration des sels solubles. Des concentrations excessives de chlorures et de sulfates peuvent causer des brûlures sur le bout des feuilles du gazon et voire même entraîner la mort des plants. Des concentrations de 250 à 400 ppm sont considérées comme indésirables pour l'irrigation des plantes sensibles aux sels.

Heureusement, les sels de chlorure et de sulfate sont rapidement solubles. Ils peuvent donc être lessivés dans les sols qui se drainent bien.

#### **4.2 Bore**

Le bore (B) est un élément mineur essentiel à la croissance de la plante mais il n'est requis qu'en minime quantité. Le bore est soluble dans l'eau et on le retrouve dans plusieurs sources d'eau utilisées pour l'irrigation. Lorsque sa concentration dans l'eau excède 1 à 2 ppm, le bore peut être toxique pour le gazon. De plus, le bore a tendance à s'accumuler dans le sol en formant des complexes chimiques qui sont difficiles à lessiver.

#### **4.3 Bicarbonates**

L'abondance des ions bicarbonates ( $\text{HCO}_3^-$ ) dans l'eau d'irrigation mérite aussi d'être évaluée. Dans le cas où l'eau d'irrigation contient un taux élevé de bicarbonates, on constate une tendance à la précipitation du calcium et du magnésium ce qui fait augmenter le SAR. On considère que l'eau d'irrigation est inadéquate pour l'irrigation lorsque la concentration résiduelle de  $\text{NaHCO}_3$  est supérieure à 2,5 mEq / L. Par contre, si la concentration est en dessous de 1,25 mEq / L, l'eau est probablement sécuritaire.

#### **5. pH de l'eau d'irrigation**

Le pH de l'eau d'irrigation devrait se situer entre 5,5 et 6,5. À ces valeurs, la solubilité de la plupart des micro-éléments est optimale. Le pH est un paramètre important en agronomie car le degré d'acidité ou de basicité de l'eau joue un rôle très important dans l'assimilation des éléments nutritifs par la plante. Il a une influence sur trois composantes importantes de la fertilité d'un sol : la biodisponibilité des nutriments, l'activité biologique et la stabilité structurale.

Lorsque le pH est inférieur à 4 il cause des dommages immédiats aux racines, et les métaux lourds comme le manganèse et le fer sont bien absorbés et peuvent empoisonner la plante. Alors que lorsque le pH est de 7 à 8, il n'est pas immédiatement dommageable. Les nutriments comme le fer, le phosphore et le manganèse sont moins disponibles ce qui causera des carences à long terme.

## Chapitre III: Régression logistique

### I- Définition de la régression logistique

La régression linéaire, simple ou multiple, est une méthode statistique très couramment utilisée dans le traitement des données, en particulier dans une démarche de modélisation.

Elle consiste à mettre en relation une variable à expliquer  $y$  avec une ou plusieurs variables explicatives  $x_1, x_2, \dots, x_p$ , appelées prédicteurs. La méthode est cependant limitée aux situations où la variable à expliquer est une variable quantitative dont la distribution, pour une valeur fixée des prédicteurs, est normale. Elle ne devrait notamment pas être utilisée lorsque la variable  $y$  est une variable qualitative. Pour de telles situations, la méthode indiquée est la régression logistique qui offre plusieurs variantes en fonction du nombre et de la nature des classes de la variable à expliquer [7].

### II- Types de régression logistique

Il y a 3 types de régression logistique selon le type de la variable à expliquer :

- **La régression logistique binaire** (binary logistic regression) : correspond au cas où la variable  $y$  comporte uniquement deux classes, les individus étant décrits par la présence ou l'absence d'un caractère donné. Par exemple, des individus (parcelles, plantes, animaux, etc.) peuvent être atteints ou non par un parasite, être fertiles ou non, être porteurs ou non d'une tare, etc.
- **La régression logistique polychotomique nominale** (polytomous nominal logistic regression) : permet de traiter les cas où la variable à expliquer possède plus de deux classes si celles-ci ne peuvent pas être ordonnées ou si on ne souhaite pas tenir compte de l'ordre dans le cas où elles seraient ordonnées. Une telle situation se présente par exemple si des individus sont caractérisés par l'appartenance à une espèce donnée, par une couleur ou par le choix d'une réponse à une question posée parmi trois propositions telles que « oui », « non », « ne sait pas ».
- Enfin, **la régression polychotomique ordinale** (polytomous ordinal regression) : concerne les situations où la variable  $y$  présente plus de deux modalités qui peuvent être ordonnées et dont on souhaite tenir compte de l'ordre. Un exemple typique est la description de l'intensité de l'attaque d'individus par un parasite, cette description étant réalisée par exemple sur la base d'une échelle à quatre niveaux notés A, B, C et D, le niveau A représentant l'absence d'attaque, le niveau B une attaque faible, le niveau C une attaque modérée et le niveau D une attaque forte [7].

### III- Intérêts de la régression logistique

À la différence de la régression linéaire (où la variable à expliquer est une variable quantitative) et du modèle de Cox (où la variable à expliquer est une variable censurée), la régression logistique s'applique lorsque la variable à expliquer (Y) est qualitative [8].

Dans le cas d'une variable explicative qualitative, une propriété très intéressante de la régression logistique est qu'elle permet d'estimer un rapport des cotes (odds ratio : OR) qui fournit une information sur la force et le sens de l'association entre la variable explicative ( $X_i$ ) et la variable à expliquer (Y). L'OR est une mesure de dépendance entre deux variables, il est toujours positif et compris entre 0 et  $+\infty$ . Lorsqu'il vaut 1, les deux variables sont indépendantes. Au contraire, plus l'OR est proche de 0 ou de  $+\infty$ , plus les variables sont liées entre elles. Si l'OR n'est pas un résultat aussi intuitif que le risque relatif, il présente l'avantage d'être utilisable quel que soit le design expérimental (cohorte ou cas-témoin). Qui plus est, si l'évènement étudiée est rare (prévalence inférieure à 5–10%), l'OR devient une bonne approximation du risque relatif. En pratique, si l'OR est supérieur à 1, on parle de facteur de risque, et si l'OR est inférieur à 1, on parle de facteur protecteur.

Dans le cas où nous souhaitons étudier plusieurs variables explicatives  $X_i$  et connaître le « poids » respectif de chacune de ces variables, un ajustement est alors nécessaire. La régression logistique est une méthode permettant de réaliser un tel ajustement. Cet ajustement consiste à individualiser « l'effet propre » de la variable explicative  $X_i$  des « effets parasites » induits par d'autres variables influençant aussi la variable à expliquer (Y) (appelées « covariables »). Cela permet ainsi de contrôler l'effet de possibles facteurs de confusion. De plus, l'ajustement permet de diminuer le bruit de fond induit par ces covariables et d'améliorer la précision de l'estimation.

Ainsi, la régression logistique tient compte de l'effet des autres variables  $X_i$  intégrées dans le modèle et permet de réaliser un ajustement de l'OR sur des covariables (on parle d'OR ajusté). La régression logistique constitue alors une méthode de choix pour rechercher et déterminer les facteurs de risque ou les facteurs protecteurs, tout en tenant compte des facteurs de confusion. Il faut cependant garder à l'esprit que l'identification d'une liaison entre une variable expliquée (Y) et une ou des variables explicatives ( $X_i$ ) ne témoigne pas forcément d'un lien de causalité et il convient de garder une certaine prudence dans l'interprétation des résultats [8].

## IV- Définition mathématique du modèle de régression logistique

Les modèles de régression sont des modèles mathématiques qui permettent de représenter une variable à expliquer (ou variable dépendante ou variable endogène ou réponse ou *outcome*) Y, en fonction d'une ou plusieurs variables  $X_1, X_2, \dots, X_p$  (dites variables *indépendantes*, variables *explicatives*, variables *exogènes*, ou *covariables*) correspondant à des facteurs de risque et de confusion potentiels. Le modèle de régression est dit *simple* s'il n'inclut qu'une seule variable explicative et *multiple* s'il inclut plusieurs variables explicatives [8].

### Quels termes introduire dans le modèle ?

#### ● Variable à expliquer

La variable à expliquer Y est la variable dont on souhaite décrire les variations en fonction d'autres variables (qui sont alors explicatives). Y peut être une variable quantitative ou qualitative. En médecine, la variable à expliquer est souvent un événement défavorable comme le décès, la maladie, la récurrence, l'erreur de prescription, etc. Mais il peut aussi s'agir d'un événement favorable (naissance, guérison, etc.) ou d'une variable quantitative (mesure biologique, score, etc.).

#### ● Variables explicatives

Les variables explicatives  $X_1, X_2, \dots, X_p$  sont des variables dont on cherche l'association avec la variable à expliquer Y ou qui peuvent modifier cette association : ce sont des facteurs de risque, des facteurs de confusion ou des facteurs d'interaction. Elles peuvent être quantitatives ou qualitatives.

#### *Facteurs de risque*

Les facteurs de risque sont des caractéristiques d'origines diverses (biologique, génétique, environnementale, culturelle, etc.) qui entraînent une augmentation de la probabilité d'apparition d'un événement défavorable.

#### *Facteurs de confusion*

Le principal intérêt de la régression multiple est de prendre en compte les biais de confusion par le principe d'ajustement. On dit qu'un facteur F joue le rôle de facteur de confusion entre l'exposition E et l'évènement M s'il est lié à la fois au facteur d'exposition E et à l'évènement M, qu'il n'est pas sur le chemin causal et s'il modifie la relation brute aux différents niveaux de F. L'association entre E et M peut être atténuée, augmentée ou inversée.

La régression logistique et la **régression linéaire** appartiennent à la même famille des modèles linéaires généraux (GLM): dans les deux cas on relie un événement à une combinaison linéaire de variables explicatives. Pour la régression linéaire, la variable dépendante suit une loi normale N ( $\mu, \sigma$ ) où  $\mu$  est une fonction linéaire des variables explicatives.

Pour la régression logistique, la variable dépendante, aussi appelée variable réponse, suit une loi de Bernoulli de paramètre  $p$  ( $p$  la probabilité moyenne pour que l'événement se produise), lorsque l'expérience est répétée une fois, ou une loi Binomiale ( $n, p$ ) si l'expérience est répétée  $n$  fois (par exemple la même dose est essayée sur  $n$  insectes). Le paramètre de probabilité  $p$  est ici une fonction d'une combinaison linéaire des variables explicatives [8].

Les fonctions les plus couramment utilisées pour relier la probabilité  $p$  aux variables explicatives sont la fonction logistique (on parle alors de modèle Logit) et la fonction de répartition de la loi normale standard (on parle alors de modèle Probit). Ces deux fonctions sont parfaitement symétriques et sigmoïdes.

La pratique de la régression logistique est très proche de celle de la régression linéaire. La régression linéaire permet de caractériser les liens entre une variable à expliquer ( $Y$ ) quantitative et des variables explicatives ( $X_1, X_2, X_3, \dots, X_p$ ) au moyen du modèle présenté par la formule :

$$Y = \beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p + \varepsilon$$

À l'évidence, ce modèle ne s'applique pas aux variables qualitatives et notamment binaires où  $Y$  s'exprime en termes de oui/non. Il est donc nécessaire d'utiliser un modèle adapté permettant de relier les variables explicatives à la variable qualitative ( $Y$ ) à prédire.

L'astuce de la régression logistique consiste non pas à modéliser la variable qualitative  $Y$  mais la probabilité que celle-ci se réalise. Le modèle logistique permet une expression non linéaire, variant de façon monotone entre 0 et 1 (Figure 2), de cette probabilité en fonction des variables explicatives ( $X_i$ ), selon la fonction sigmoïde suivante [8] :

$$P(Y = 1 / X) = \pi(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

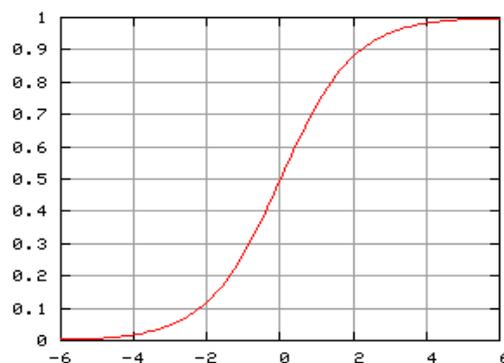


Figure 2 : Exemple de modèle de régression logistique.

Propriétés intéressantes :

$$X \rightarrow +\infty \quad \text{alors } \pi(X) \rightarrow 1$$

$$X \rightarrow -\infty \quad \text{alors } \pi(X) \rightarrow 0$$

$$\pi(X) \in [0 ; 1]$$

#### IV-1 Modèle logistique à une variable explicative X

La variable à expliquer Y est la survenue ou non d'un événement. Cette variable est binaire : présence (+) ou absence (-). Considérons une seule variable explicative (X) (régression logistique simple). Le modèle est présenté par la formule :

$$Y = \pi (X) + \varepsilon = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} + \varepsilon$$

Nous voyons dans cette formule que l'OR lié à la variable d'exposition (X) correspond à l'exponentielle du coefficient  $\beta_1$ . Le coefficient  $\beta_1$  correspond donc au logarithme de l'odds ratio ( $\beta_1 = \log (OR)$ ) mesurant l'association entre la variable explicative (X) et la variable expliquée (Y). Dans le cas d'une telle régression logistique simple, la fonction logistique a une forme de sigmoïde [9].

#### IV-2 Modèle logistique à plusieurs variables explicatives X

Un modèle multivarié (régression logistique multiple) permet d'exprimer une variable Y en fonction de plusieurs variables  $X_i$ .

L'écriture du modèle est très similaire à celle présenté par le modèle logistique simple. À chaque variable explicative,  $X_i$ , est associé un coefficient  $\beta_i$  et donc un  $OR_i$  correspondant là aussi à l'exponentielle de ce coefficient et qui mesure l'association entre la variable  $X_i$  et la maladie (Y) en question. Pour chaque variable explicative, ce coefficient est ajusté sur les autres variables explicatives du modèle [9].

$$Y = P (Y= 1 / \{X_j\}) + \varepsilon = \pi (\{X_j\}) + \varepsilon = \frac{e^{\beta_0 + \sum \beta_j X_j}}{1 + e^{\beta_0 + \sum \beta_j X_j}} + \varepsilon$$

#### IV-3 Transformation et régression LOGIT

Dans un modèle de régression logit, les valeurs prévues de la variable dépendante ou de réponse ne seront jamais inférieures ou égales à 0 ni supérieures ou égales à 1, et ce quelles que soient les valeurs des variables indépendantes ; par conséquent, il est courant d'analyser des variables dépendantes ou de réponse binaires. Ceci peut être accompli en appliquant l'équation de régression suivante (le terme *logit* a été utilisé pour la première fois par Berkson, 1944) [9] :

$$Y = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$

On voit aisément que, indépendamment des coefficients de régression ou de l'importance des valeurs de x, ce modèle produira toujours des valeurs prévues comprises entre 0 et 1.

Le nom **logit** fait référence au fait qu'il est possible de linéariser simplement ce modèle en appliquant une transformation logarithmique.

Si nous pensons à une variable dépendante binaire Y en termes de probabilité continue sous-jacente  $\pi$ , comprise entre 0 et 1, nous pouvons transformer cette probabilité  $\pi$  en :

$$\pi' = \log\left(\frac{\pi}{1 - \pi}\right)$$

Cette transformation est appelée transformation **logit ou logistique**. Notez que théoriquement,  $\pi'$  peut prendre toutes valeurs comprises entre  $-\infty$  et  $+\infty$ . Comme la transformation logit résout le problème des bornes 0/1 de la variable dépendante originale (probabilité), nous pouvons utiliser ces valeurs (*transformées logit*) dans une équation de régression linéaire ordinaire. En fait, si nous appliquons la transformation logit des deux côtés de l'équation de régression logit donnée plus tôt, nous obtenons le modèle standard de **régression multiple** linéaire [9] :

$$\pi' = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \times X_1 + \dots + \beta_p \times X_p$$

#### **IV-4 Sélection des variables explicatives à inclure dans le modèle de régression logistique**

Le choix des variables explicatives ( $X_i$ ) n'est pas le fruit du hasard ni le résultat d'un screening fait à l'aveugle de plusieurs centaines de variables. Il est basé sur la connaissance du phénomène étudié et les possibles facteurs influençant. Ainsi, toutes les variables explicatives étudiées ne seront pas toutes nécessairement incluses dans l'analyse multivariée. Ne seront introduites dans la régression logistique que les variables qui pourraient avoir un lien avec la variable expliquée [8].

Le choix des variables est donc fondé sur leur pertinence clinique et sur la connaissance de facteurs de confusion avérés ou supposés. Cette première étape permettra de choisir les variables explicatives ( $X_i$ ) les plus pertinentes à étudier.

La deuxième étape consiste à décrire de manière précise ces variables explicatives, à en vérifier l'exhaustivité (absence de données manquantes) et la qualité (absence de biais de classification). Il est souvent difficile de se faire une idée sur cette étape fondamentale à la vue d'une publication. Ensuite est réalisée une analyse univariée : cette étape consiste donc à estimer l'association entre la variable (Y) et chaque variable  $X_i$  en réalisant une régression logistique simple pour chaque variable  $X_i$ . Les OR ainsi calculés sont dits « bruts » ou « non ajustés ». Les variables explicatives qui sont liées de façon suffisamment forte à la variable à expliquer sont alors conservées dans le modèle. D'une manière générale, toutes les variables dont le degré de significativité est inférieur à 0,20 (c'est à dire avec une valeur de probabilité  $< 0,20$ ) lors de l'analyse univariée seront incluses dans le modèle initial de régression logistique multiple.

Ce seuil de 0,20, et non pas 0,05 comme habituellement utilisé en statistique, permet de prendre en compte des variables qui pourraient être des facteurs de confusion possible ou des facteurs d'interaction.

Bien entendu un tel seuil peut sembler arbitraire et peut varier selon les habitudes des différentes équipes (il est parfois de 0,25, parfois de 0,30). Il est important de

Signaler que seront aussi incluses dans l'analyse des variables connues pour être associées à la réponse, même si l'analyse univariée n'a pas abouti à une valeur de probabilité  $< 0,20$ . On parle de variables dites « forcées ».

Lors de cette démarche, il ne faut pas inclure des variables redondantes (trop liées entre elles) dans l'analyse de régression logistique multiple. Dans un tel cas, on parle de colinéarité entre ces variables. Ce problème a pour conséquence un modèle de qualité médiocre, voire une impossibilité d'estimation des coefficients. La colinéarité entre les variables doit donc être vérifiée lors de la réalisation d'une régression logistique. Cela suppose d'étudier les liens existant entre les variables explicatives [8].

#### **IV-5 Obtention du modèle final**

Lorsqu'il s'agit d'ajuster une comparaison sur différentes variables prédéfinies, les variables explicatives  $X_i$  sont choisis a priori et aucune procédure de choix n'est nécessaire. En revanche, lorsque le but est le développement d'un modèle explicatif, certaines procédures peuvent être mises en œuvre pour sélectionner les variables à garder dans le modèle final. Ces procédures ont pour but de vérifier le principe de parcimonie, qui en science et en philosophie, consiste à n'utiliser que le minimum de causes élémentaires pour expliquer un phénomène. En d'autres termes, l'objectif est de sélectionner le modèle apportant le maximum d'information sur la variable à expliquer (Y) à partir du plus petit nombre de variables explicatives possible ( $X_i$ ).

Au-delà de l'intérêt épistémologique, cela facilite grandement les estimations du modèle et cela limite les problèmes liés aux données manquantes.

Sur un plan pratique, il n'y a pas de règle unanime concernant la méthode à utiliser. Parmi les méthodes possibles, il est souvent proposé des procédures pas à pas ou **stepwise**, plus ou moins automatiques. Avec ces méthodes, le choix du modèle final se fait de manière progressive. La procédure en pas à pas ascendante ou **forward selection** consiste à inclure progressivement les variables explicatives ( $X_i$ ) à un modèle minimaliste en laissant de côté celles qui n'apportent pas suffisamment d'information au modèle. La procédure en pas à pas descendante ou **backward elimination** est la plus couramment utilisée.

Elle consiste à inclure toutes les variables sélectionnées au préalable et à retirer progressivement celles qui n'apportent pas suffisamment d'information au modèle. Avec ces méthodes, on sélectionne le modèle en laissant de côté les variables qui ne sont pas significatives (là aussi le niveau de significativité n'est pas forcément fixé à 0,05 et peut être fixé à 0,10, 0,15, ou 0,20 par exemple).

Une fois le modèle final obtenu, on vérifie s'il existe des interactions entre les variables explicatives ( $X_i$ ) du modèle. L'existence d'une interaction entre deux variables signifie que l'effet d'une variable est différent en fonction de l'autre variable. Il faut souligner que ces étapes de sélection d'un modèle demandent beaucoup de sens critique, et ne doivent surtout pas être basées seulement sur le raisonnement statistique. En effet, certaines variables jugées indispensables peuvent être forcées dans le modèle.

Enfin, il peut être parfois proposé plusieurs modèles explicatifs en intégrant certaines variables ou non. Cela peut se faire, par exemple, quand leur impact est controversé ou encore pour simplifier le modèle. Il peut s'agir d'un argument scientifique supplémentaire montrant la robustesse de l'association des variables d'intérêt avec la variable à expliquer  $Y$  au travers de plusieurs modèles [8].

#### **IV-6 Estimation des paramètres par la maximisation de la vraisemblance**

##### **a) Principe de la vraisemblance**

L'estimation des paramètres pour la régression linéaire multiple se fait par la méthode des moindres carrés ordinaires. Pour la régression logistique, la méthode des moindres carrés ne permet pas d'obtenir une estimation des coefficients. On utilise donc la méthode du maximum de vraisemblance (Maximum likelihood). C'est une méthode classique qui permet d'estimer les paramètres d'une loi, d'un modèle [9].

Pour estimer les paramètres de la régression logistique par la méthode du maximum de vraisemblance, nous devons tout d'abord déterminer la loi de distribution de  $P(Y=X)$ .

$Y$  est une variable binaire définie dans  $\{+; -\}$  ou  $\{1; 0\}$  pour simplifier les écritures Pour un individu  $i$  ; on modélise la probabilité à l'aide de la loi binomiale  $B(1; \pi)$ , avec :

$$P[Y(i) = X(i)] = \pi(i)^{y(i)} \times (1 - \pi(i))^{(1 - y(i))}$$

Cette modélisation est cohérente avec ce qui a été dit précédemment, en effet :

- Si  $Y(i) = 1$ , alors  $P[Y(i) = 1 = X(i)]$
- Si  $Y(i) = 0$ , alors  $P[Y(i) = 0 = X(i)]$

- **Vraisemblance**

La vraisemblance (en anglais likelihood) d'un échantillon  $\Omega$  s'écrit :

$$L = \prod \pi(i)^{y(i)} \times (1 - \pi(i))^{(1-y(i))}$$

Pour alléger l'écriture, nous utiliserons pour la suite :

$$L = \prod \pi^y \times (1 - \pi)^{(1-y)}$$

N'oublions pas que la vraisemblance correspond à la probabilité d'obtenir l'échantillon  $\Omega$  à partir d'un tirage dans la population. Elle varie donc entre 0 et 1. La méthode du maximum de vraisemblance consiste à produire les paramètres  $a = (a_0; a_1; \dots a_J)$  de la régression logistique qui rendent maximum la probabilité d'observer cet échantillon.

- **Log-vraisemblance**

Pour faciliter les manipulations, on préfère souvent travailler sur la log-vraisemblance (log-likelihood) :

$$LL = \sum_i y \times \ln \pi + (1 - y) \times \ln (1-\pi)$$

Le logarithme étant une fonction monotone, le vecteur  $\mathbf{a}$  qui maximise la vraisemblance est le même que celui qui maximise la log-vraisemblance. Cette dernière varie entre  $-\infty$  et 0.

Puisque  $\hat{a}$  est un estimateur du maximum de vraisemblance, il en possède toutes les propriétés :

1. Il est asymptotiquement sans biais ;
2. Il est de variance minimale ;
3. Il est asymptotiquement gaussien.

Ces éléments, notamment le dernier, seront très importants pour l'inférence statistique (intervalle de confiance, test de significativité, etc.).

- **Déviante**

Bien souvent, on utilise la quantité :

$$DM = -2 LL = -2 \times LL$$

Appelée déviance (ou déviance résiduelle, en anglais résiduel déviance, dans certains logiciels tels que R) (DM). Contrairement à la log-vraisemblance, elle est positive. L'objectif de l'algorithme d'optimisation est de minimiser cette déviance. On peut faire le parallèle avec la somme des carrés des résidus de la régression linéaire multiple. La déviance nulle (D0) calculée sur le modèle uniquement composée de la constante correspondrait alors à la somme des carrés totaux :

$$D = 2 \times \ln \times \frac{L(\text{modèle saturé})}{L(\text{modèle étudié})}$$

$$\begin{aligned}
&= -2 \times LL (\text{Modèle étudié}) - [-2 \times LL (\text{Modèle saturé})] \\
&= DM - [-2 \times LL (\text{Modèle saturé})] \\
&= -2 \sum [-y \times \ln \left( \frac{\hat{\pi}}{y} + (1 - y) * \ln \left( \frac{1 - \hat{\pi}}{1 - y} \right) \right)]
\end{aligned}$$

Un modèle saturé pour des données individuelles est un modèle reconstituant parfaitement les valeurs de la variable dépendante c.-à-d.  $\hat{\pi}(i) = y(i)$ . Sa vraisemblance est égale à 1, et sa log-vraisemblance 0. Dans ce contexte,  $D = DM$  [9].

### b) Test global de significativité

Le test de significativité du modèle de la régression logistique se fait par le test de rapport de vraisemblance. C'est un test d'hypothèse qui évalue l'adéquation de l'ajustement d'un modèle afin de déterminer le meilleur ajustement pour les données échantillons : un modèle non contraint dont tous les paramètres sont libres et son modèle contraint correspondant avec moins de paramètres pour tester l'hypothèse nulle [9].

- Hypothèses du test de rapport de vraisemblance

$H_0$  : Pas de liaison entre Y et les  $X_j \rightarrow \beta_1 = \beta_2 = \dots = \beta_p = 0$

$H_1$  : Le modèle a du sens  $\rightarrow$  Au moins 1  $\beta_j \neq 0$

### **b- 1 : Extension du test à une seule variable explicative X**

- Principe du test de rapport de vraisemblance (TRV)

Le principe est basé sur la comparaison de la vraisemblance  $L_X$  (avec variable explicative (H1)) avec la vraisemblance  $L_0$  sans variable explicative (H0).

$\rightarrow$  Intuitivement si  $L_X > L_0$  alors la variable X apporte à l'estimation de P(Y)

### **Construction de la statistique de test**

$L_X$  vraisemblance avec X  $\rightarrow$  déjà calculée

$L_0$  vraisemblance sans X (sous H0)

Sur l'échantillon de taille  $N$  on observe :  $\text{Card} \{y = 1\} = n_1$  et  $\text{Card} \{y = 0\} = N - n_1$

Avec  $\hat{\pi} = P(Y = 1) = n_1 / N$  et  $1 - \hat{\pi} = P(Y = 0) = 1 - n_1 / N$

Aussi  $L_0 = \prod \hat{\pi}^{y_i} \times (1 - \hat{\pi})^{(1-y_i)} = \prod (n_1 / N)^{y_i} \times (1 - (n_1 / N))^{(1-y_i)}$

$$= (n_1 / N)^{\sum y_i} \times (1 - (n_1 / N))^{\sum (1-y_i)}$$

$$L_0 = (n_1 / N)^{n_1} \times (1 - (n_1 / N))^{N - n_1}$$

- Statistique de test du TRV

On montre que sous  $H_0$  :  $D = -2 \ln (L_0/L_X) \sim \chi^2_{1 \text{ ddl}}$

- Interprétation

→ Non rejet de  $H_0$  : le modèle n'a pas de sens, X n'apporte rien à l'estimation de  $P(Y)$ .

→ Rejet de  $H_0$  : le modèle a du sens, X apporte à l'estimation de  $P(Y)$

## **b-2 : Extension du test à p variables explicatives $X_j$ :**

- Hypothèse du test

$H_0 : \beta_1 = \beta_2 = \dots \beta_p = 0$  : absence de liaison

$H_1$  : Il existe au moins un  $\beta_j \neq 0$  : présence de liaison

- Statistique du test sous

$$D = -2 \ln (L_0/L_X) \sim \chi^2_{p \text{ ddl}}$$

Si on ne rejette pas  $H_0$  associée au TRV alors on s'arrête.

Si, par contre, on rejette  $H_0$  alors on réalise un test individuel de chaque coefficient

- Test sur un coefficient – Hypothèses

$H_0 : \beta_j = 0$  (la variable n'est pas significative dans le modèle)

$H_1 : \beta_j \neq 0$  (la variable est significative dans le modèle)

- Statistique du test (Test de Wald)

On peut montrer que si  $H_0$  est vraie alors :

$$K = (j^{\wedge} / S^2_{\beta^{\wedge} j}) \sim \chi^2_{1 \text{ ddl}}$$

## **V - Domaines d'application de la régression logistique**

La régression logistique est largement répandue dans de nombreux domaines. On peut citer de façon non exhaustive :

- En médecine, elle permet par exemple de trouver les facteurs qui caractérisent un groupe de sujets malades par rapport à des sujets sains.
- Dans le domaine des assurances, elle permet de cibler une fraction de la clientèle qui sera sensible à une police d'assurance sur tel ou tel risque particulier.
- Dans le domaine bancaire, pour détecter les groupes à risque lors de la souscription d'un crédit.
- En économétrie, pour expliquer une variable discrète. Par exemple, les intentions de vote aux élections.

## **PARTIE 2 : Matériel et méthodes**

### **CHAPITRE I : Etude expérimentale**

#### **I- Description de la zone d'étude**

Skhirate-Témara est une préfecture qui appartient à la région de Rabat-Salé Kénitra et dont le chef-lieu est Témara. Créée en 1983, elle s'étend sur un territoire de 485 km<sup>2</sup>. Elle est limitée par l'oued Ykem au Nord-Est, l'oued Cherrât au Sud et l'océan atlantique à l'Ouest.

Le climat de cette région est méditerranéen de par sa douceur tant hivernale qu'estivale et l'irrégularité de ses précipitations. Il est, ainsi, tempéré, doux et humide en raison de l'influence atlantique le long de la côte (25 km de plages).

Skhirat est caractérisée par de fortes activités maraîchères basées sur l'irrigation par les eaux souterraines. La surexploitation de ces ressources en eau conduit au problème de salinisation des eaux et de pollution nitrique ce qui rend l'eau de mauvaise qualité. Pour la résolution de ce problème on fait l'étude de la qualité des eaux d'irrigation dans cette région. Pour ce faire, un suivi des paramètres physico-chimiques des eaux d'irrigation a été réalisé pour évaluer leur qualité.

#### **II- Méthodes d'échantillonnage de l'eau**

##### **II-1- Choix des sites de prélèvement**

Après avoir défini la zone d'étude sur une carte topographique à échelle de 1/50 000, les points d'échantillonnage ont été identifiés. Cela a été réalisé après des visites avec des représentants du "Centre des Travaux (CT)" de Skhirat, une subdivision administrative locale du ministère de l'Agriculture.

Ainsi, 71 puits creusés à la main ont été choisis pour suivre la caractérisation de qualité de l'eau (Dakak et al, 2014). Les emplacements spécifiques des puits étudiés ont été pris avec l'aide d'un récepteur GPS.

Le choix des sites d'échantillonnage reposait principalement sur deux critères :

- les caractéristiques spatiales des différentes zones de la région de Skhirat
- les niveaux de salinité de l'eau d'irrigation.

##### **II -2 Méthode de prélèvement**

Des échantillons d'eau ont été prélevés dans des bouteilles en polyéthylène de 1 litre qui ont été pré-nettoyées avec de l'acide chlorhydrique concentré et de l'eau distillée.

Certains paramètres ont été mesurés in situ, tels que le pH mesuré à l'aide d'un pH-mètre portable, la conductivité électrique (CE) mesuré à l'aide d'un conductimètre portable et le niveau piézométrique au moyen d'une sonde.

### **III- Analyse des paramètres physico-chimiques**

En plus des paramètres mesurés in situ, différentes autres analyses physico-chimiques de l'eau ont été menées au laboratoire de chimie des sols de l'Unité de Recherche sur l'Environnement et la Conservation des Ressources Naturelles de l'INRA à Rabat (Zouahri et al, 2015).

#### **III-1 Analyse des paramètres physico-chimiques de l'eau in situ**

##### **III-1-1 pH**

Le pH permet de déterminer l'acidité de l'eau. Il a été mesuré in situ à l'aide d'un pH mètre.

##### **III-1-2 Conductivité électrique (CE)**

La mesure de la conductivité électrique des eaux a été réalisée sur place à l'aide d'un conductimètre. Elle est exprimée en dS/m.

##### **III-1-3 Niveau piézométrique**

Afin de connaître l'état de la ressource en eau en termes de quantité, il est indispensable de surveiller la profondeur de la surface des nappes (niveau piézométrique).

Elle peut être mesurée ponctuellement par une sonde manuelle, ou régulièrement à l'aide d'enregistreur automatique, avec données télétransmises pour le plus perfectionné. Les mesures sont réalisées dans des forages, puits ou piézomètres.

La piézométrie est la mesure de profondeur de la surface de la nappe d'eau souterraine. Elle est exprimée par rapport au sol en m.

#### **III-2 Analyse des paramètres physico-chimiques de l'eau au laboratoire**

##### **Bilan ionique**

##### **III-2-1 Dosage du sodium ( $\text{Na}^+$ ) et du potassium ( $\text{K}^+$ )**

Le dosage des ions de  $\text{Na}^+$  et  $\text{K}^+$  se fait par un appareil de spectrophotomètre à flamme. Le principe de dosage de ces 2 ions est basé sur la dissociation de  $\text{Na}^+$  et  $\text{K}^+$  contenus dans la solution à doser pendant leur passage dans une flamme de gaz butane, ces derniers émettent de l'énergie à leur état d'excitation. Chaque élément va émettre une longueur d'onde bien précise qui sera détectée par des filtres optiques spécifiques.

Le principe consiste à réaliser une courbe d'étalonnage de l'élément à doser avec des concentrations connues, puis évaporer dans la flamme du photomètre l'eau à analyser et déterminer la teneur du Na<sup>+</sup> ou K<sup>+</sup> en connaissant la longueur d'onde d'absorption de chacun d'eux.

### III-2-2 Dosage du calcium (Ca<sup>2+</sup>) et du magnésium (Mg<sup>2+</sup>)

Le dosage des ions calcium et magnésium est un dosage complexométrique.

Dans une fiole de 100 ml, on met 10 ml de l'échantillon d'eau et on complète jusqu'à 100 ml par l'eau distillée.

On verse la solution dans un erlenmeyer, et on ajoute 5 gouttes de KCl, et 1 ml de triéthanolamine à 97%.

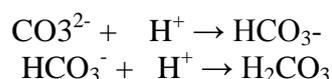
- Pour le dosage des ions Ca<sup>2+</sup>, on ajoute 20 ml de la soude NaOH et quelques gouttes d'un indicateur coloré, et on dose par l'EDTA à 0.02N tout en agitant la solution.
- Alors pour le magnésium, on ajoute 20 ml de la solution tampon et quelques gouttes d'un autre indicateur coloré et on dose par l'EDTA.

### III-2-3 Dosage des carbonates (CO<sub>3</sub><sup>2-</sup>) et bicarbonates (HCO<sub>3</sub><sup>-</sup>)

Le dosage du carbonate et du bicarbonate est un dosage acido-basique. Dans un bécher, on met 10 ml de l'échantillon d'eau à analyser et on ajoute quelques gouttes de phénolphtaléine.

- L'apparition d'une coloration rose indique la présence des ions CO<sub>3</sub><sup>2-</sup> dans l'échantillon à analyser. On dose ensuite par HCl à 0.05N. L'absence de la couleur rose indique l'absence des ions carbonates.
- Concernant les ions bicarbonates HCO<sub>3</sub><sup>-</sup>, on procède de la même manière, seulement au lieu de phénolphtaléine on ajoute quelques gouttes de méthylorange.

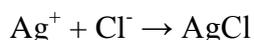
Les réactions de ce dosage sont :



### II2-4 Dosage des chlorures (Cl<sup>-</sup>)

Les chlorures ont été déterminés par la méthode de Mohr. Cette méthode est basée sur un dosage par précipitation. Dans un bécher, on met l'échantillon d'eau à analyser, et on ajoute trois gouttes de chromate de potassium K<sub>2</sub>CrO<sub>4</sub>.

On dose la solution par le nitrate d'argent AgNO<sub>3</sub> (0.02N) tout en agitant. A chaque quantité d'ion argent ajouté, se forme un précipité de nitrate d'argent :



### III-3-5 Dosage des sulfates ( $\text{SO}_4^{2-}$ )

Les sulfates ont été dosés par la méthode néphélométrique qui consiste à précipiter les sulfates dont on prélève 10 ml de l'échantillon d'eau dans un erlenmeyer de 250 ml, on verse ainsi un volume V ml de HCl nécessaire pour doser les carbonates et les bicarbonates, on porte le mélange à l'ébullition pendant 5 minutes puis on ajoute goutte à goutte 10 ml exactement mesurés de  $\text{BaCl}_2$ , après on fait bouillir pendant 5 minutes et on laisse refroidir. Juste avant le titrage, on ajoute 2 ml de  $\text{MgCl}_2$ , puis 20 ml de tampon ( $\text{NH}_4\text{OH}-\text{NH}_4\text{Cl}$ ), le titrage est réalisé par E.D.T.A en présence de noir Eriochrome T jusqu'à virage du violet au bleu franc.

#### IV - Normes de qualité des eaux destinées à l'irrigation

Ces normes sont reprises dans le tableau 2 [10].

Tableau 2 : Normes indicatives de la qualité de l'eau d'irrigation [10].

Paramètres	Très bonne qualité	Qualité moyenne	Mauvaise qualité
SAR	$SAR < 10$	$10 < SAR < 18$	$18 < SAR < 26$
Conductivité électrique (CE) en dS/m	CE < 0,25 Non saline	$0.25 < CE < 2,25$ Salinité moyenne	$2,25 < EC < 5$ EC > 5 Salinité élevée
Risque lié aux chlorures (Cl-) en méq/l	$0,14 < Cl < 0,85$ eau extrêmement fraîche	$0,85 < Cl < 8,46$ eau douce	$8,46 < Cl < 282,06$ eau saumâtre
pH	pH < 4	$5,5 < pH < 6,5$	pH > 8
Sulfates (SO <sub>4</sub> <sup>2-</sup> ) en méq/l	$SO_4^{2-} < 250$	$250 < SO_4^{2-} < 400$	$SO_4^{2-} > 400$
Sels solubles totaux (SST) en g/l	SST < 0.3	$0.45 < SST < 4.6$	SST > 5
Risque lié aux carbonates et aux bicarbonates CO <sub>3</sub> <sup>2-</sup> HCO <sub>3</sub> <sup>-</sup> (RCB) en méq/l	RCB < 1.25	$1.25 < RCB < 2.5$	RCB > 2.5

## Chapitre II : Logiciel SPSS

### I- Présentation du logiciel SPSS

**SPSS** (*Statistical Package for the Social Sciences*) est un logiciel utilisé pour l'analyse statistique. C'est aussi le nom de la société qui le revend (SPSS Inc). En 2009, la compagnie décide de changer le nom de ses produits en **PASW** pour *Predictive Analytics Software*.

La première version de SPSS a été mise en vente en 1968 et fait partie des programmes utilisés pour l'analyse statistique en sciences sociales. Il est utilisé par des chercheurs en économie, en science de la santé, par des compagnies d'études, par le gouvernement, des chercheurs de l'éducation nationale, etc. En plus de l'analyse statistique, la gestion des données (sélection de cas, reformatage de fichier, création de données dérivées) et la documentation des données (un dictionnaire de métadonnées est sauvegardé avec les données) sont deux autres caractéristiques du logiciel.

SPSS impose des contraintes concernant la structure interne des fichiers, les types des données, le traitement des données et les fichiers de correspondance. Les séries de données de SPSS ont une structure en tableau à deux dimensions où les rangées représentent typiquement les cas (tels des individus ou des foyers) et où les colonnes représentent les mesures (telles que âge, sexe, ou revenu du foyer). Il n'y a que deux types définis de données : nombres ou chaînes de caractères. Tous les traitements de données se déroulent de façon séquentielle les cas à la suite les uns des autres dans l'ordre du fichier. Les fichiers peuvent être comparés un à un, un envers les autres, mais pas plusieurs d'un coup envers les autres [11].

### II- Méthodes de régression logistique dans SPSS

Les méthodes de régression logistique disponibles sont les mêmes que pour la régression linéaire. Toutefois, le critère de sélection pour les méthodes progressives est différent.

Vous pouvez donc opter pour la méthode Entrée et insérer toutes les variables prédictives en même temps. De plus, si vous préférez sélectionner l'ordre d'entrée des variables, choisissez la méthode hiérarchique. Les paramètres seront calculés pour chaque bloc de variables.

Parmi les méthodes progressives, vous avez toujours le choix entre ascendante ou descendante. Dans la méthode ascendante, SPSS introduit la variable ayant le score le plus élevé en premier jusqu'à ce qu'aucune variable n'ait une statistique (score) significative (soit plus petit que 0,05). Dans la méthode descendante, le contraire se produit puisque le premier modèle évalué contient toutes les variables et SPSS retire celles qui ne contribuent pas significativement à l'amélioration de la prédiction.

La différence avec la régression linéaire multiple est que SPSS évalue à chaque étape si certaines variables devraient être retirées en se basant sur différents critères :

\* Le **rapport de vraisemblance** (*likelihood-ratio*, LR) : SPSS conserve la variable si le changement du LR est significatif quand la variable est retirée, ce qui indique que cette variable contribue à la qualité de l'ajustement

\*La **statistique conditionnelle** : il s'agit d'un critère moins exigeant que le LR, donc il est préférable de prioriser le 1<sup>er</sup>

\*La **statistique Wald** : cette fois, SPSS retire toutes les variables pour lesquelles la statistique Wald est inférieure à 0,1. Cette méthode peut être utilisée avec un petit échantillon. Sinon, il est préférable de privilégier le LR [11].

## PARTIE 3 : Résultats et discussion

### Chapitre I : Statistique descriptive

#### I - Statistique descriptive

La statistique descriptive a pour objectif de résumer l'information contenue dans un échantillon. Elle permet de décrire la façon dont les résultats des observations se distribuent en utilisant différentes paramètres numériques telle que les paramètres de position (la moyenne), les paramètres de dispersion (le minimum, le maximum et l'écart-type) ainsi que le coefficient de variation. Ces paramètres sont repris dans le Tableau 3.

La statistique descriptive permet de montrer aussi la différence entre la valeur maximale et la valeur minimale pour toutes les variables. A partir des résultats qu'on a trouvé on constate que les variables CE, pH, K<sup>+</sup>, HCO<sub>3</sub><sup>-</sup>, et CO<sub>3</sub><sup>2-</sup> ont des amplitudes (E=max-min) petites c.à.d. qu'il n'y a pas une grande différence entre le maximum et le minimum de ces variables. Les variables Ca<sup>2+</sup>, Mg<sup>2+</sup>, Na<sup>+</sup>, Cl<sup>-</sup> et SO<sub>4</sub><sup>2-</sup> ont des amplitudes plus ou moins élevées que les autres variables. La variable NO<sub>3</sub><sup>-</sup> est caractérisé par une amplitude très élevé.

A partir des résultats de la distribution des données, on remarque bien que les variables CE, pH, HCO<sub>3</sub><sup>2-</sup>, K<sup>+</sup>, et CO<sub>3</sub><sup>2-</sup> ont une faible dispersion des données autour de la moyenne (écart-type très faible), contrairement aux variables Ca<sup>2+</sup>, Mg<sup>2+</sup>, SO<sub>4</sub><sup>2-</sup>, Na<sup>+</sup> et Cl<sup>-</sup> qui ont une dispersion moyenne des données, suivies par la variables NO<sub>3</sub><sup>3-</sup> qui est caractérisée par une grande dispersion (écart-type très élevé).

Les valeurs élevées du coefficient de variation pour les variables Mg<sup>2+</sup>, K<sup>+</sup> et NO<sub>3</sub><sup>-</sup> montrent la grande dispersion de ces valeurs alors pour le reste des paramètres ont un CV faible à moyen donc leur dispersion est faible.

Tableau 3: Paramètres statistiques descriptifs de l'eau d'irrigation.

	CE	Ph	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Na <sup>+</sup>	K <sup>+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>--</sup>	HCO <sub>3</sub> <sup>-</sup>	CO <sub>3</sub> <sup>--</sup>	NO <sub>3</sub> <sup>-</sup>
Min	0.90	6.20	0.48	0.10	13.68	0.01	2.72	0.00	1.35	0.00	0.00
Max	8.08	8.24	26.80	27.80	76.43	1.09	63.40	32.02	9.73	2.05	256.65
Moyenne	3.17	7.45	7.41	4.98	26.45	0.15	21.21	8.25	5.09	0.84	50.61
Ecart-type	1.72	0.31	6.20	5.30	11.75	0.16	15.00	6.95	1.58	0.52	54.44
CV (%)	54.24	4.12	83.64	106.43	44.42	112.43	70.69	84.23	31.11	62.12	107.57

## II- Etude de la corrélation entre tous les paramètres.

Selon les résultats de la corrélation de Pearson (Tableau 4), on déduit qu'il y a une corrélation très significative entre plusieurs variables:

- L'existence d'une corrélation forte et positive entre la variable CE et les variables  $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Na^+$ ,  $K^+$ ,  $Cl^-$  et  $SO_4^{2-}$   
Fortement et positivement corrélée avec la variable  $Na^+$   
Fortement et négativement corrélée avec la variable  $Ca^{2+}$   
Moyennement et positivement corrélée avec la variable  $CO_3^{2-}$   
Moyennement et négativement corrélée avec la variable  $Mg^{2+}$
- Par ailleurs la variable pH est fortement et négativement corrélée avec les variables CE,  $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Cl^-$  et  $SO_4^{2-}$ .
- L'existence d'une corrélation forte et positive entre les variables CE,  $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Na^+$ ,  $K^+$ ,  $Cl^-$  et  $SO_4^{2-}$ .
- La variable  $NO_3^-$  présente une corrélation forte et négative avec la variable  $CO_3^{2-}$

Pour la régression logistique on s'intéresse à l'étude de deux paramètres de la qualité de l'eau d'irrigation qui sont le SAR et la CE et leurs Co-variables, parce que ce sont les paramètres responsables des risques les plus dangereux pour l'irrigation puisque le SAR est responsable du risque d'alcalinité et la CE est responsable du risque de salinité.

Tableau 4. Coefficients de corrélation de Pearson entre les différents paramètres chimiques.

V	CE	pH	Ca2+	Mg2+	Na+	K+	Cl-	SO4--	HCO3-	CO3--	SAR
CE	1,00	<b>-0,32</b>	<b>0,72</b>	<b>0,73</b>	<b>0,91</b>	<b>0,47</b>	<b>0,98</b>	<b>0,83</b>	0,07	0,06	0,09
		0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,57	0,62	0,47
pH	<b>-0,32</b>	1,00	<b>-0,38</b>	<b>-0,35</b>	-0,18	<b>-0,10</b>	<b>-0,32</b>	<b>-0,36</b>	0,21	0,03	0,19
		0,01	0,00	0,00	0,13	0,43	0,01	0,00	0,08	0,80	0,12
Ca2+	<b>0,72</b>	<b>-0,38</b>	1,00	<b>0,62</b>	<b>0,43</b>	0,23	<b>0,77</b>	<b>0,64</b>	-0,13	-0,19	<b>-0,48</b>
		0,00		0,00	0,00	0,06	0,00	0,00	0,27	0,11	0,00
Mg2+	<b>0,73</b>	<b>-0,35</b>	<b>0,62</b>	1,00	<b>0,64</b>	<b>0,36</b>	<b>0,74</b>	<b>0,83</b>	0,07	-0,03	<b>-0,26</b>
		0,00	0,00		0,00	0,00	0,00	0,00	0,59	0,83	0,03
Na+	<b>0,92</b>	-0,18	<b>0,43</b>	<b>0,64</b>	1,00	<b>0,53</b>	<b>0,88</b>	<b>0,78</b>	0,16	0,14	<b>0,36</b>
		0,00	0,13	0,00	0,00		0,00	0,00	0,18	0,24	0,00
K+	<b>0,47</b>	-0,10	0,23	<b>0,36</b>	<b>0,53</b>	1,00	<b>0,49</b>	<b>0,36</b>	0,07	-0,03	0,14
		0,00	0,43	0,06	0,00	0,00		0,00	0,54	0,84	0,26
Cl-	<b>0,98</b>	<b>-0,32</b>	<b>0,77</b>	<b>0,74</b>	<b>0,88</b>	<b>0,49</b>	1,00	<b>0,78</b>	-0,04	0,02	0,03
		0,00	0,00	0,00	0,00	0,00		0,00	0,75	0,88	0,78
SO4--	<b>0,83</b>	<b>-0,36</b>	<b>0,64</b>	<b>0,83</b>	<b>0,78</b>	<b>0,36</b>	<b>0,78</b>	1,00	0,10	-0,07	-0,07
		0,00	0,00	0,00	0,00	0,00	0,00		0,43	0,57	0,57
HCO3-	0,07	0,21	-0,13	0,07	0,16	0,07	-0,04	0,10	1,00	0,21	0,20
		0,57	0,08	0,27	0,59	0,18	0,54	0,75	0,43		0,08
CO3--	0,06	0,03	-0,19	-0,03	0,14	-0,03	0,02	-0,07	0,21	1,00	<b>0,27</b>
		0,62	0,80	0,11	0,83	0,24	0,84	0,88	0,57	0,08	0,03
SAR	0,09	0,19	<b>-0,48</b>	<b>-0,26</b>	<b>0,36</b>	0,14	0,03	-0,07	0,20	<b>0,27</b>	1,00
		0,47	0,12	0,00	0,03	0,00	0,26	0,78	0,57	0,10	0,03

\*\* . La corrélation est significative au niveau 0.01 (bilatéral).

\* . La corrélation est significative au niveau 0.05 (bilatéral).

Première ligne : coefficient de corrélation de Pearson

Deuxième ligne : probabilité  $\alpha_{observée}$  indiquant le degré de significativité du test.

### III- Etude de la corrélation entre le SAR et les autres paramètres

D'après le Tableau 4, ci-dessus, le SAR est significativement corrélé avec  $Ca^{2+}$  (-0.48),  $Mg^{2+}$  (-0.26),  $Na^+$  (0.36) et  $CO_3^{2-}$  (0.27). La relation entre le SAR et ces 4 variables est illustrée sur les figures 3 à 6.

Le nuage de points du SAR et  $Ca^{2+}$  augmente de gauche à droite et de bas en haut (Figure 3). Une droite de régression se rapprochant le plus de tous les points peut être tracée afin de montrer la force et la direction de la relation entre les 2 variables.

Le SAR et  $Ca^{2+}$  évoluent en sens inverse ; la bande de points va de la gauche supérieure à la droite inférieure parce que la corrélation entre SAR et  $Ca^{2+}$  est moyenne et négative.

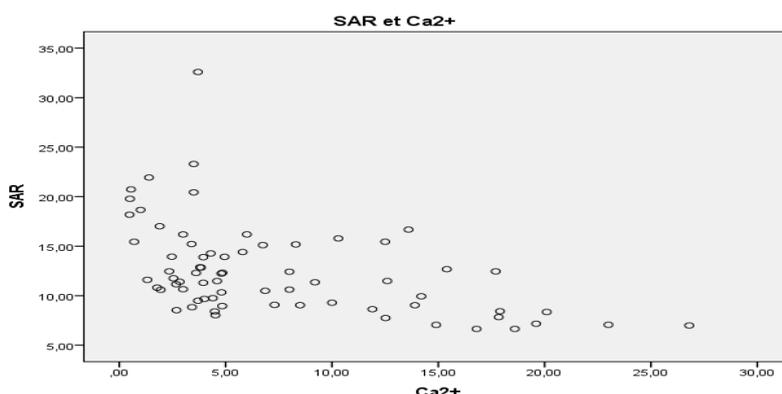


Figure 3. Diagramme de dispersion SAR-Ca<sup>2+</sup>.

Le nuage des points du SAR et  $Mg^{2+}$  augmente de gauche à droite et de bas en haut (Figure 4). Une droite de régression se rapprochant le plus de tous les points peut être tracée afin de montrer la force et la direction de la relation entre les 2 variables. Le SAR et  $Mg^{2+}$  évoluent en sens inverse ; la bande de points va de la gauche supérieure à la droite inférieure parce que la corrélation entre SAR et  $Mg^{2+}$  est négative.

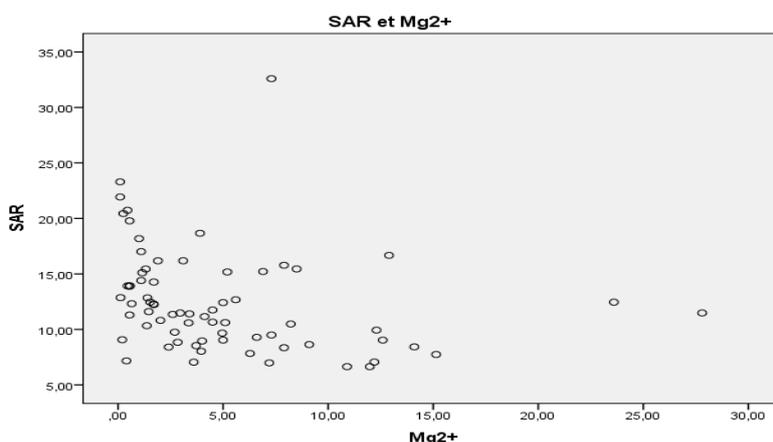


Figure 4. Diagramme de dispersion SAR-Mg<sup>2+</sup>.

Le nuage des points du SAR et  $\text{Na}^+$  montre l'accroissement simultané des deux paramètres (Figure 5). Le SAR et  $\text{Na}^+$  évoluent dans le même sens; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre SAR et  $\text{Na}^+$  est positive.

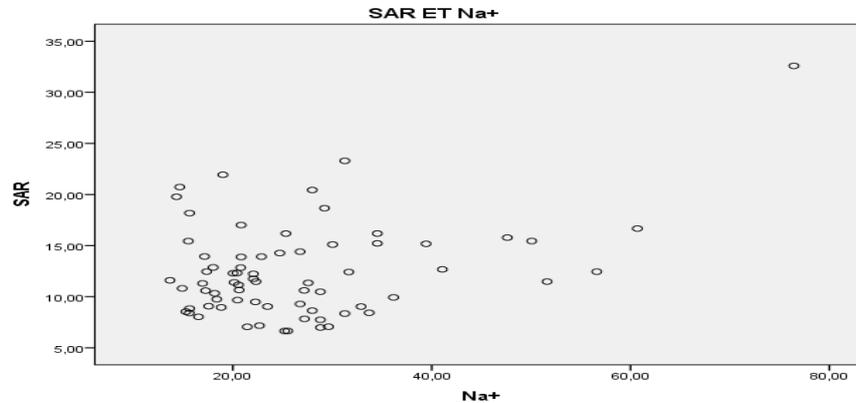


Figure 5. Diagramme de dispersion SAR- $\text{Na}^+$ .

Le nuage des points du SAR et  $\text{CO}_3^{2-}$  (Figure 6) ne présente pas une relation forte entre ces deux paramètres, il montre que la corrélation positive entre ces 2 variables a une faible intensité. La direction de la relation entre les 2 variables évolue dans le même sens.

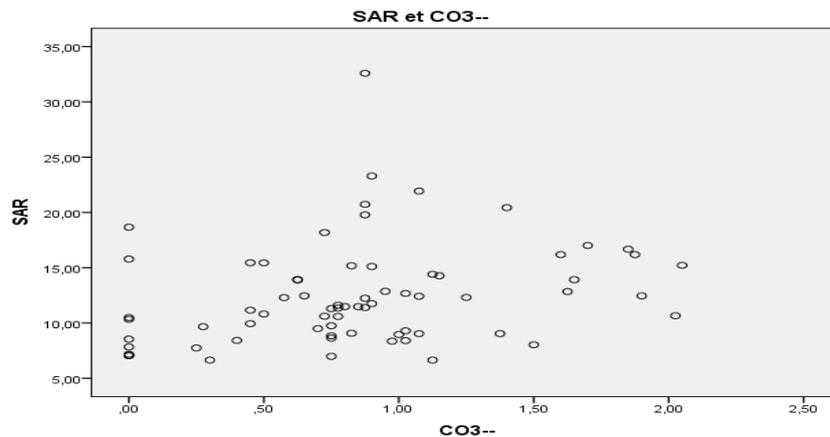


Figure 6. Diagramme de dispersion SAR- $\text{CO}_3^{2-}$ .

#### IV- Etude de la corrélation entre la CE et les autres paramètres

Le Tableau 4 montre que la CE est significativement corrélée avec  $\text{Ca}^{2+}$  (0.72), pH (-0.32),  $\text{Mg}^{2+}$  (0.73),  $\text{Na}^+$  (0.92),  $\text{K}^+$  (0.47),  $\text{Cl}^-$  (0.98) et  $\text{SO}_4^{2-}$  (0.83). Ces relations sont indiquées sur les figures 7 à 13.

Le nuage des points de CE et pH augmente de droite à gauche et de bas en haut (Figure 7). Une droite de régression se rapprochant le plus de tous les points peut être tracée afin de montrer la force et la direction de la relation entre les 2 variables.

CE et pH évoluent en sens inverse ; la bande de points va de la gauche supérieure à la droite inférieure parce que la corrélation entre CE et pH est négative.

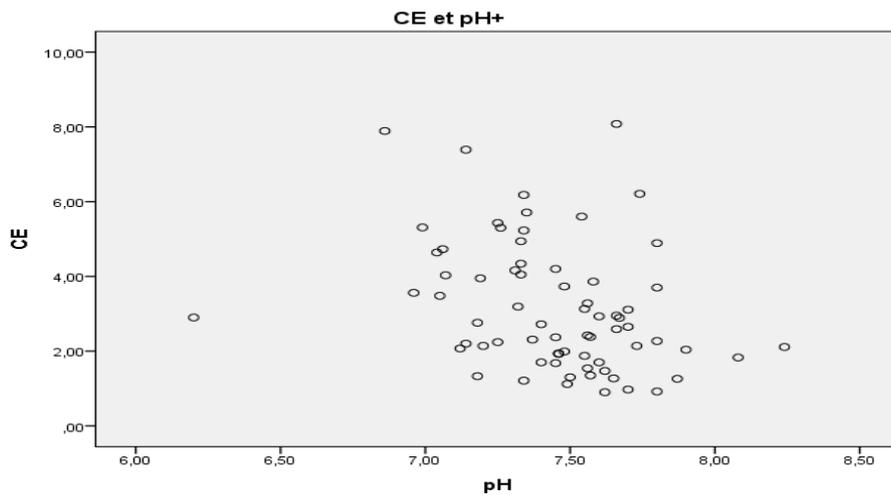


Figure 7. Diagramme de dispersion CE-pH.

Le nuage de points de CE et  $\text{Ca}^{2+}$  augmente de gauche à droite et de bas en haut (Figure 8).

Une droite de régression se rapprochant le plus de tous les points peut être tracée afin de montrer la force et la direction de la relation entre les 2 variables.

CE et  $\text{Ca}^{2+}$  évoluent dans le même sens ; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre CE et  $\text{Ca}^{2+}$  est positive.

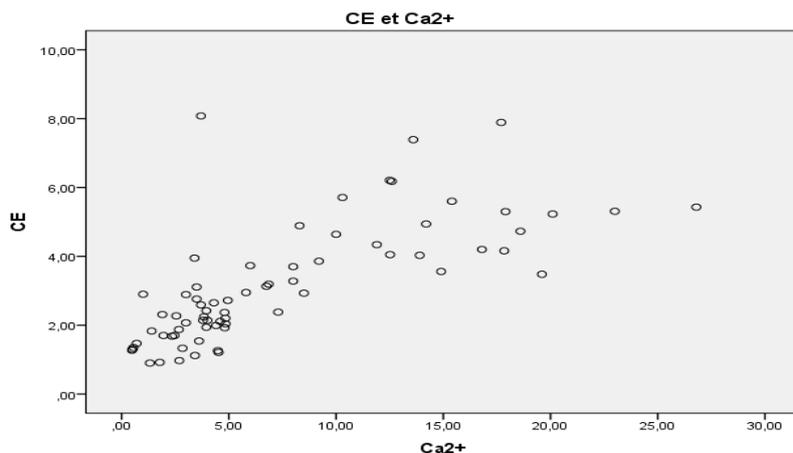


Figure 8. Diagramme de dispersion CE-Ca2+.

Le nuage de point de CE et  $\text{Mg}^{2+}$  augmente de gauche à droite et de bas en haut (Figure 9).

Une droite de régression se rapprochant le plus de tous les points peut être tracée afin de montrer la force et la direction de la relation entre les 2 variables. La CE et  $\text{Mg}^{2+}$  évoluent dans le même sens ; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre CE et  $\text{Mg}^{2+}$  est positive.

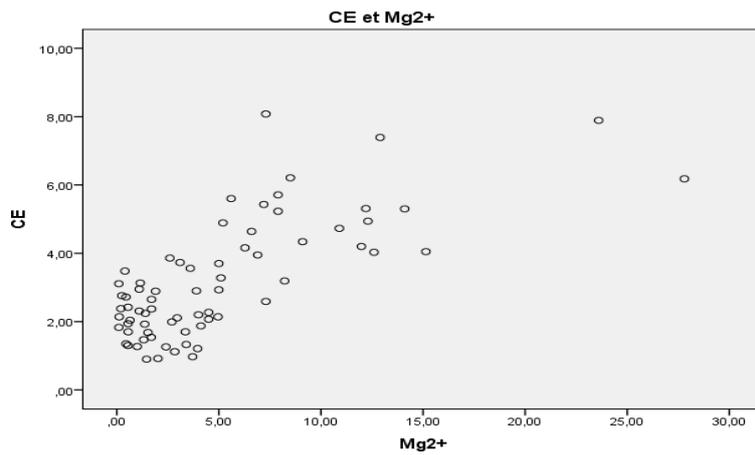


Figure 9. Diagramme de dispersion CE-Mg<sup>2+</sup>.

Le nuage de points de CE et Na<sup>+</sup> montre l'accroissement simultané des deux paramètres (Figure 10). La CE et Na<sup>+</sup> évoluent dans le même sens; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre SAR et Na<sup>+</sup> est positive.

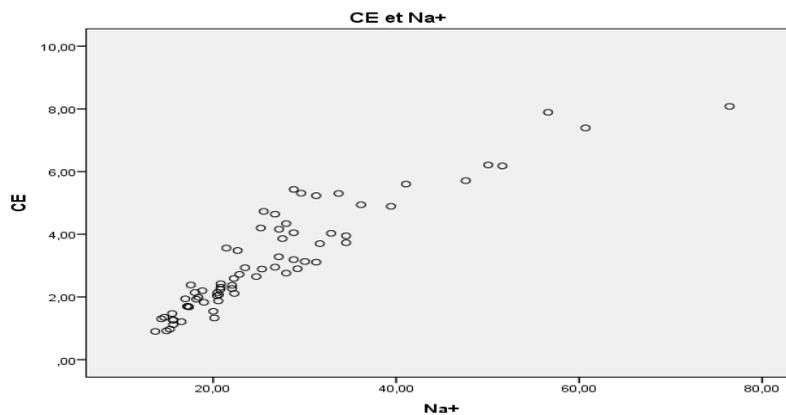


Figure 10. Diagramme de dispersion CE-Na<sup>+</sup>.

Le nuage des points de CE et K<sup>+</sup> augmente de gauche à droite et de bas en haut (Figure 11). Une droite de régression se rapprochant le plus de tous les points peut être tracée afin de montrer la force et la direction de la relation entre les 2 variables. La CE et K<sup>+</sup> évoluent dans le même sens; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre CE et K<sup>+</sup> est positive.

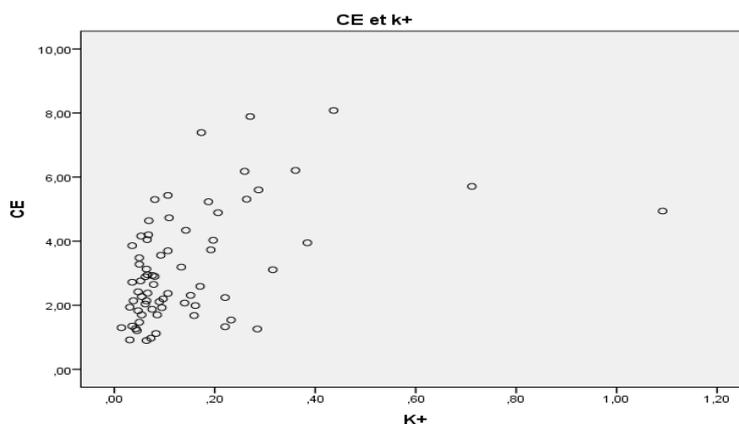


Figure 11. Diagramme de dispersion CE-K<sup>+</sup>.

Le nuage des points de CE et Cl<sup>-</sup> montre l'accroissement simultané des deux paramètres (Figure 12). La CE et Cl<sup>-</sup> évoluent dans le même sens; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre CE et Cl<sup>-</sup> est forte et positive.

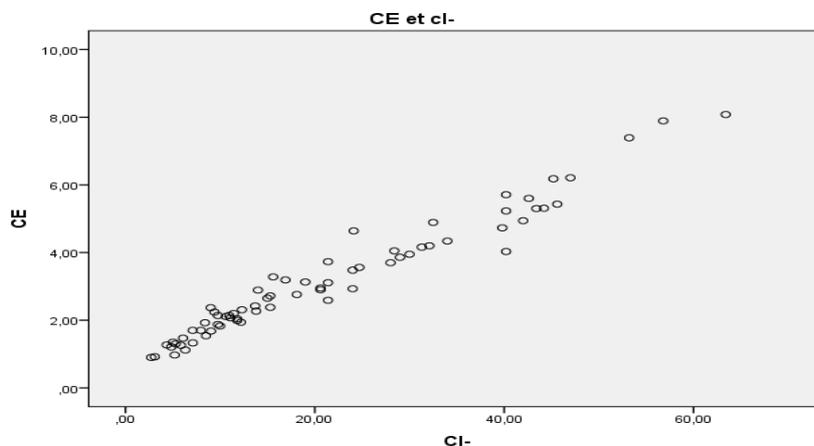


Figure 12. Diagramme de dispersion CE-Cl-.

Le nuage des points de CE et SO<sub>4</sub><sup>2-</sup> montre l'accroissement simultané des deux paramètres (Figure 13). La CE et SO<sub>4</sub><sup>2-</sup> évoluent dans le même sens; la bande de points va de la gauche inférieure à la droite supérieure parce que la corrélation entre CE et SO<sub>4</sub><sup>2-</sup> est positive.

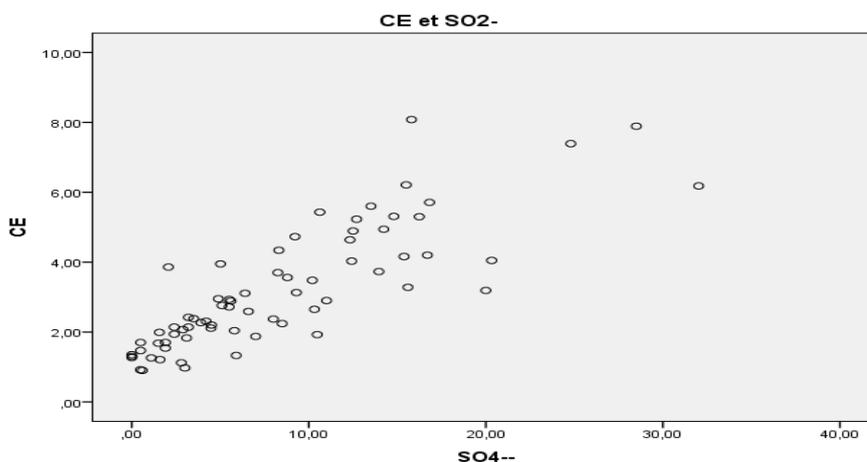


Figure 13. Diagramme de dispersion CE-SO4--.

## Chapitre II : Régression logistique

SPSS nous permet d'avoir différentes étapes dans notre modèle de régression logistique. La différence entre les étapes c'est l'inclusion des prédicteurs.

Ceci est similaire à bloquer des variables dans des groupes et ensuite les entrer dans l'équation un groupe à la fois. Par défaut, la régression logistique SPSS est exécutée en deux étapes. La première étape, appelée bloc 0, n'inclut pas de prédicteurs mais seulement l'ordonnée à l'origine (la constante). Souvent, ce modèle n'est pas intéressant pour les chercheurs.

La seconde étape, bloc 1 - C'est la première étape (ou modèle) avec des prédicteurs. Dans ce cas, c'est le modèle complet que nous avons spécifié dans la commande de régression logistique. Nous pouvons avoir plus d'étapes si on procède par étapes ou si on utilise le blocage des variables.

### I- Régression logistique binaire pour le SAR

#### Évaluation de la signification du modèle de régression

Dans le tableau relatif aux tests de spécification du modèle, nous constatons pour la 3<sup>ème</sup> étape (la dernière) que l'étape et le modèle complet sont significatifs.

La ligne étape montre en effet la différence entre la probabilité -2LL de l'étape précédente et celle obtenue par l'ajout du nouveau prédicteur.

La ligne bloc ne sera examinée que dans une régression hiérarchique où on introduirait plus d'une variable (donc un bloc de variables) par étape.

Tableau 5. Tableau des tests de spécification du modèle de régression SAR<sub>binaire</sub>.

		Khi-Chi-deux	Ddl	Sig.
Etape 3 <sup>a</sup>	Etape	-17.093	1	.000
	Bloc	32.660	2	.000
	Modèle	32.660	2	.000

Nous pouvons ensuite examiner le test de Hosmer-Lemeshow. Celui-ci indique s'il existe un écart important entre les valeurs prédites et observées. Nous constatons à la lecture du tableau qu'à l'étape 3 il n'existe pas une différence significative (valeur  $p = 0.994 \gg 0,05$ ) entre les valeurs prédites et observées. Ça veut dire que le modèle a une bonne qualité prédictive.

Tableau 6. Tableau de test de Hosmer-Lemeshow pour modèle de régression SAR<sub>binaire</sub>.

Etape	Khi-Chi-deux	Ddl	Sig.
3	1.434	8	.994

### Évaluation de l'ajustement des données au modèle de régression

Ce tableau présente les valeurs de l'équation de régression logistique pour prédire la variable dépendante (SAR binaire) à partir des variables indépendantes. Ils sont dans des unités log-odds.

Dans les 2 étapes précédentes on a trouvé que les paramètres sont tous non significatifs, mais dans la 3<sup>ème</sup> étape après l'élimination des 2 paramètres  $Mg^{2+}$  et  $CO^{3-}$  les résultats sont devenus significatifs.

L'équation de régression logistique est :  $SAR_{\text{binaire}} = 3.956 + 2.717 * Ca^{2+} - 4.19 * Na^{+}$

Le test Wald est pour évaluer la signification statistique des coefficients estimés des variables indépendantes conservées afin de s'assurer que chacune contribue à mieux prédire P (SAR binaire = 1).

Exp (B) - Ce sont les rapports de cotes pour les prédicteurs. Ils sont l'exponentiation des coefficients. Donc tous les paramètres sont entrés dans l'équation de régression logistique.

Tableau 7. Tableau des variables dans l'équation pour modèle de régression SAR<sub>binaire</sub>.

		A	E.S.	Wald	Ddl	Sig.	Exp(B)
Etape 3	Ca2+	2.717	1.109	6.002	1	.014	15.132
	Na+	-.419	.198	4.469	1	.035	.658
	Constante	3.956	2.528	2.448	1	.118	52.249

### Évaluation de la qualité d'ajustement du modèle

Ce tableau montre combien de cas de la variable dépendante (SAR) sont correctement prédits basés sur le modèle de régression logistique complet (6 cas observés sont 0 et sont correctement prédits à 0, et 60 cas sont observés à 1 et sont correctement prédits à 1) alors que seulement 4 observations qui sont incorrectes.

Ce modèle logistique complet du SAR binaire a une bonne prédiction des variables avec un pourcentage de 94.3%.

Comme vous pouvez le voir, ce pourcentage est passé de 88,6% pour le modèle de base à 94.3% pour le modèle complet du SAR binaire.

Tableau 8. Tableau de classement pour modèle de régression SAR<sub>binaire</sub>.

		Observations		Prévisions		
				sar bin		Pourcentage correct
		0	1	0	1	
Etape 1	sar bin	0		6	2	75.0
		1		2	60	96.8
	Pourcentage global					

- **Calcul de la sensibilité et de la spécificité du modèle de SAR binaire**

La sensibilité désigne la probabilité d'obtenir un test positif sur un échantillon négatif (le faux négatif). Au contraire, la spécificité désigne la probabilité d'obtenir un test négatif sur un échantillon positif (le faux positif). Le concept de sensibilité et de spécificité est utilisé pour les tests dichotomiques (oui/non, positif/négatif, etc.)

Sensibilité :  $Se = 60 / (60+2) * 100 = 96.77\%$ .

Une sensibilité de 0.967 signifie que, lorsque l'échantillon est positif, il y a 96.77% de chance que la prédiction du modèle de SAR binaire soit positive.

Spécificité :  $Sp = 6 / (6+2) * 100 = 75\%$

Une spécificité de 0,75 signifie que, lorsque l'échantillon est négatif, il y a 75% de chance que la prédiction du modèle de SAR binaire soit négative.

## **II- Régression logistique ordinale pour le SAR**

Nous voyons ce tableau qui donne la  $-\log$  de la vraisemblance pour le modèle ayant seulement l'ordonnée à l'origine (la constante) et du modèle final. La vraisemblance du modèle est utilisée pour tester si tous les coefficients de régression estimés dans le modèle sont simultanément zéro.

Chi-carré : C'est le test du chi-carré du rapport de vraisemblance. Il teste si au moins un des coefficients de régression des prédicteurs n'est pas égal à zéro dans le modèle.

L'hypothèse nulle stipule que tous les coefficients de régression dans le modèle sont égaux à zéro. Si elle est rejeté nous amènerait à conclure qu'au moins un des coefficients de régression dans le modèle n'est pas égal à zéro.

Tableau 9. Tableau des informations d'ajustement de modèle de régression SAR<sub>ordinal</sub>.

Modèle	-2log-vraisemblance	Chi-Square	Ddl	Sig.
Seulement la constante	138,543			
Final	65,556	72,988	4	,000

## Évaluation de l'ajustement des données au modèle de régression

Dans ce tableau, nous voyons les coefficients, leurs erreurs-types, le test de Wald et les p-valeurs associées (Sig.), ainsi que l'intervalle de confiance à 95% des coefficients.

Les 3 variables  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  et  $\text{Na}^+$  sont statistiquement significatives, donc chacune d'elles contribue à mieux prédire la réponse P (SAR ordinaire) alors que  $\text{CO}_3^{2-}$  est non significative donc elle ne contribue pas à la prédiction du modèle SAR ordinaire.

Pour une augmentation d'une unité de  $\text{Na}^+$ , nous nous attendons à une augmentation de 0,345 des probabilités logarithmiques ordonnées d'être à un niveau supérieur. toutes les autres variables du

modèle sont maintenues constantes.

Pour une augmentation d'une unité de  $\text{Ca}^{2+}$ , nous nous attendons à une augmentation de -0.785 des probabilités logarithmiques ordonnées d'être à un niveau inférieur, toutes les autres variables du modèle sont maintenues constantes. La même chose est valable pour les autres variables.

Dans ce cas les ordonnées à l'origine (constantes) seraient 1.369, 6.728 et 15.619.

L'équation de régression logistique est :

$$\text{SAR}_{\text{ordinal}} = 1.369 \text{ (pour SAR ordinal = 1) ou } 6.728 \text{ (pour SAR ordinal = 2) ou } + 15.61 \text{ - (pour SAR ordinal = 3) } + 0.785 * \text{Ca}^{2+} - 0.229 * \text{Mg}^{2+} + 0.345 * \text{Na}^+$$

Tableau 10. Tableau des estimations de paramètre du modèle régression SAR<sub>ordinal</sub>.

		Estimation	Erreur std.	Wald	Ddl	Sig.	Intervalle de confiance 95 %	
							Borne inférieure	Borne supérieure
Seuil	[SARord = 1]	1,369	1,029	1,771	1	,183	-,647	3,386
	[SARord = 2]	6,728	1,596	17,771	1	,000	3,600	9,856
	[SARord = 3]	15,619	14,892	1,100	1	,294	-13,568	44,806
Emplacement	Ca2	-,785	,184	18,256	1	,000	-1,145	-,425
	Mg2	-,229	,100	5,232	1	,022	-,426	-,033
	Na	,345	,077	20,026	1	,000	,194	,496
	CO3	-,131	,631	,043	1	,835	-1,367	1,105

## Évaluation de la qualité de l'ajustement du modèle

Il est maintenant possible d'examiner si le modèle permet de bien classer les sujets dans leur groupe d'appartenance à partir de l'équation finale. Nous voyons que le pourcentage correct de classification est :

$$(18+34+2+1)/70 * 100 = 57.01 \%$$

Alors ce modèle de SAR ordinal est de moyenne qualité de prédiction de résultats.

Donc on conclut que l'application d'un modèle de SAR binaire qui a une bonne prédiction des résultats (94.3%) est beaucoup mieux que l'application du modèle de SAR ordinal qui a une qualité de prédiction moyenne.

Tableau 11. Tableau croisé SAR ord \* Modalité de réponse prévue.

			Modalité de réponse prévue				Total
			1	2	3	4	
SAR ord	1	Effectif	18	6	0	0	24
		Effectif théorique	6,9	15,4	1,4	,3	24,0
	2	Effectif	2	34	2	0	38
		Effectif théorique	10,9	24,4	2,2	,5	38,0
	3	Effectif	0	5	2	0	7
		Effectif théorique	2,0	4,5	,4	,1	7,0
	4	Effectif	0	0	0	1	1
		Effectif théorique	,3	,6	,1	,0	1,0
Total	Effectif	20	45	4	1	70	
	Effectif théorique	20,0	45,0	4,0	1,0	70,0	

## III- Régression logistique binaire pour la CE

### Évaluation de la signification du modèle de régression

A l'étape 1, nous constatons que l'étape (soit l'ajout de la variable) et le modèle complet sont significatifs. Bien sûr, à l'étape 1, le modèle ne comprend qu'une variable CI, donc nécessairement, la valeur  $\chi^2$  est identique pour les deux éléments. La ligne bloc ne sera examinée que dans une régression hiérarchique où on introduirait plus d'une variable (donc un bloc de variables) par étape.

À la lumière de ce tableau, nous pouvons dire que le modèle final permet de prédire significativement mieux la probabilité de CE binaire.

Tableau 12. Tableau des tests de spécification du modèle de base de CE<sub>binaire</sub>.

		Khi-Chi-deux	Ddl	Sig.
Etape 1	Etape	77,781	1	,000
	Bloc	77,781	1	,000
	Modèle	77,781	1	,000

Nous pouvons ensuite examiner le test de Hosmer-Lemeshow. Celui-ci indique s'il existe un écart important entre les valeurs prédites et observées. Nous constatons à la lecture du tableau qu'il n'existe pas d'une différence significative (valeur p 0.400 >> 0.05) entre les valeurs prédites et observées. Ça veut dire que le modèle a une très bonne qualité prédictive.

Tableau 13. Tableau de test de Hosmer-Lemeshow pour le modèle de base de CE<sub>binaire</sub>.

Etape	Khi-Chi-deux	Ddl	Sig.
1	8,355	8	0,400

### Évaluation de l'ajustement des données au modèle de régression

Ce tableau présente les valeurs de l'équation de régression logistique pour prédire la variable dépendante (CE binaire) à partir des variables indépendantes. Ils sont dans des unités log-odds.

L'équation de régression logistique est :

$$CE_{\text{binaire}} = 13.963 - 1.129 * CI$$

Tableau 14. Tableau des variables dans l'équation de modèle de base de CE<sub>binaire</sub>.

		A	E.S.	Wald	Ddl	Sig.	Exp(B)
Etape 1 <sup>a</sup>	CI-	-1,129	,419	7,249	1	,007	,323
	Constante	13,963	5,077	7,564	1	,006	1159074,023

### Évaluation de la qualité d'ajustement du modèle

Ce tableau montre combien de cas de la variable dépendante (CE) sont correctement prédits basés sur le modèle de régression logistique complet (41 cas observés sont 0 et sont correctement prédits à 0 et 27 cas sont observés à 1 et sont correctement prédits à 1) alors que seulement 2 observations qui sont incorrectes.

Ce modèle logistique complet de la CE binaire a une bonne prédiction de la variable avec un pourcentage de 97.1%.

Comme vous pouvez le voir, ce pourcentage est passé de 61,4% pour le modèle de base à 97.1% pour le modèle complet de la CE binaire.

Tableau 15. Tableau de classement pour le modèle de régression de CE<sub>binaire</sub>.

		Observations		Prévisions		
				CE bin		Pourcentage correct
		0	1			
Etape 1	CE bin	0	41	2	95,3	
		1	0	27	100,0	
	Pourcentage global				97,1	

## Calcul de la sensibilité et de la spécificité du modèle de CE binaire

Sensibilité :  $Se = 27 / (27+2) * 100 = 93.1\%$

Une sensibilité de 0.931 signifie que, lorsque l'échantillon est positif, il y a 93.1% de chance que la prédiction du modèle de CE binaire soit positive.

Spécificité :  $Sp = 41 / (41+0) * 100 = 100\%$

Une spécificité de 1 signifie que, lorsque l'échantillon est négatif, il y a 100% de chance que la prédiction du modèle de CE binaire soit négatif.

## **IV- Régression logistique ordinal pour la CE**

Nous voyons ce tableau qui donne la  $-2\log$  de la vraisemblance pour le modèle de l'ordonnée à l'origine (constante) seulement et du modèle final. La vraisemblance du modèle est utilisée pour tester si tous les coefficients de régression estimés dans le modèle sont simultanément zéro.

Chi-carré : C'est le test du chi-carré du rapport de vraisemblance. Il teste si au moins un des coefficients de régression des prédicteurs n'est pas égal à zéro dans le modèle.

L'hypothèse nulle suppose que tous les coefficients de régression dans le modèle sont égaux à zéro. Si elle est rejetée, on conclut qu'au moins un des coefficients de régression dans le modèle n'est pas égal à zéro.

Tableau 16. Tableau des informations d'ajustement de modèle de régression de CE<sub>ordinal</sub>.

Modèle	-2log-vraisemblance	Chi-Square	ddl	Sig.
Seulement la constante	135,037			
Final	,000	135,037	2	,000

## Évaluation de l'ajustement des données au modèle de régression

Les 2 variables  $Ca^{2+}$  et  $Na^{+}$  sont statistiquement significatives, donc chacune d'elles contribue à mieux prédire la réponse  $P(y)$ .

Dans ce cas les ordonnées à l'origine (constantes) seraient 48.625 pour CE ordinaire = 1 et 75.249 pour CE ordinaire = 2.

L'équation de régression logistique est :

$CE_{ordinal} = 48.625$  (pour CE ordinaire = 1) ou  $75.249$  (pour CE ordinaire = 2)  $+1.675 * Ca^{2+} + 1.399 * Na^{+}$

Tableau 17. Tableau des estimations des paramètres pour le modèle de régression de CE<sub>ordinal</sub>.

	Estimation	Erreur std.	Wald	ddl	Sig.	Intervalle de confiance 95 %		
						Borne inférieure	Borne supérieure	
Seuil	[CEord = 1]	48,625	24,702	3,875	1	,049	,210	97,040
	[CEord = 2]	75,249	37,742	3,975	1	,046	1,276	149,222
Emplacement	Ca2	1,675	,829	4,086	1	,043	,051	3,300
	Na	1,399	,732	3,651	1	,056	-,036	2,834

## Évaluation de la justesse de l'ajustement du modèle

Il est maintenant possible d'examiner si le modèle permet de bien classer les sujets dans leur groupe d'appartenance à partir de l'équation finale. Nous voyons que le pourcentage correct de classification est :

$$(40+19+11) / 70 * 100 = 100\%$$

Alors ce modèle de CE ordinal a une excellente et parfaite qualité prédictive.

Tableau 18. Tableau croisé CE ord \* prédiction CE<sub>ordinal</sub>.

			prédiction CE ord			Total
			1	2	3	
CE ord	1	Effectif	40	0	0	40
		Effectif théorique	22,9	10,9	6,3	40,0
	2	Effectif	0	19	0	19
		Effectif théorique	10,9	5,2	3,0	19,0
	3	Effectif	0	0	11	11
		Effectif théorique	6,3	3,0	1,7	11,0
Total		Effectif	40	19	11	70
		Effectif théorique	40,0	19,0	11,0	70,0

Donc on conclut que l'application d'un modèle de régression logistique ordinal à la CE a une très bonne qualité prédictive (100%), et beaucoup mieux que l'application du modèle de régression logistique binaire.

## *Conclusions*

L'eau d'irrigation a une influence sur la qualité du sol et des cultures. Par conséquent, l'objectif de ce travail était l'évaluation de la qualité de l'eau d'irrigation dans la région de Skhirat.

Cette évaluation est faite par la méthode de la régression logistique. La surveillance des paramètres physico-chimiques de l'eau a été réalisée dans 71 puits. Des paramètres tels que le pH et la conductivité électrique ont été mesurés in situ tandis que d'autres comme l'équilibre ionique ont été mesurés en laboratoire alors que d'autres paramètres ont été calculés comme le SAR.

Le traitement des données par la régression logistique montre que les eaux souterraines dans la région présentent un risque moyen de salinité avec un faible risque d'alcalinité.

Après l'analyse des données par nos modèles de régressions on conclut que les modèles de prédiction de la CE par l'utilisation de la régression logistique binaire et ordinaire et du SAR par l'utilisation de la régression logistique binaire ont une très bonne qualité prédictive des résultats alors que le modèle de régression logistique ordinaire pour le SAR a une qualité prédictive moyenne.

En résumé, bien que l'eau souterraine dans la région de Skhirat présente un moyen risque de salinité, elle est de bonne qualité pour l'irrigation. Les pratiques agricoles devraient être bien gérées pour sécuriser l'utilisation de la ressource en eau pour un développement durable de l'agriculture dans la région.

## *Références bibliographiques*

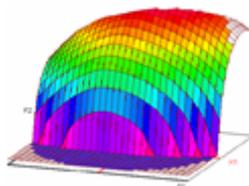
- [1] [www.inra.org.ma/rabat](http://www.inra.org.ma/rabat)
- [2] Rapport d'activité 2011 de CRRRA de Rabat
- [3] [www.lenntech.fr](http://www.lenntech.fr) - L'eau d'irrigation –Lenntech
- [4] AgriMaroc: <http://www.agrimaroc.ma/les-differentes-techniques-d-irrigation/>
- [5] Bureau d'Etudes Industrielles "Energies Renouvelables et Environnement" 2006
- [6] Dakak H, Zouahri A, Iaaich H, Moussadek R, El Khadir M, Douaik A, Soudi B et Benmohammadi A. (2014). Apports des systèmes d'information géographique au diagnostic de la pollution nitrique des eaux souterraines : cas de la zone de Skhirate, Maroc. *Revue des Régions Arides*, 33 : 75-79
- [7] Cour de principaux modèles utilisés en régression logistique. | Université de Liège
- [8] Cour Comprendre la régression logistique. EM| consulte
- [9] Cour de régression logistique- EA 2694 - Santé Publique : Epidémiologie et Qualité des soins - Université de Lille 2
- [10] Zouahri A, Dakak H, Douaik A, El Khadir M et Moussadek R. (2015). Evaluation of the groundwater suitability for irrigation in the Skhirate region, Northwest of Morocco. *Environmental Monitoring and Assessment*, 187: 4184.
- [11] <https://developer.ibm.com/predictiveanalytics/2017/08/08/spss-statistics-25-now-available-purchase>

# Annexes

Les résultats des analyses des paramètres physico-chimiques de 70 échantillons d'eau prélevés à partir des puits dans la zone de Skhirat.

X_COORD	Y_COORD	CE	pH	Ca2+	Mg2+	Na+	K+	Cl-	SO4--	HCO3-	CO3--	NO3-	SAR
-6.97	33.82	2.04	7.90	4.85	0.65	20.43	0.06	11.80	5.80	6.70	1.25	0.00	12.32
-6.98	33.82	3.48	7.05	19.60	0.40	22.67	0.05	24.00	10.20	6.53	0.00	81.78	7.17
-6.98	33.83	5.43	7.25	26.80	7.20	28.81	0.11	45.60	10.62	3.70	0.75	66.12	6.99
-6.99	33.83	3.56	6.96	14.90	3.60	21.45	0.09	24.70	8.79	5.55	0.00	12.62	7.05
-6.99	33.83	1.87	7.55	2.67	4.12	20.56	0.08	9.76	7.00	4.50	0.45	43.50	11.16
-6.98	33.84	2.59	7.66	3.70	7.30	22.26	0.17	21.40	6.59	3.75	0.70	54.81	9.49
-6.97	33.85	5.31	6.99	23.00	12.20	29.62	0.26	44.20	14.80	4.03	0.00	65.25	7.06
-7.01	33.83	2.93	7.60	8.50	5.00	23.49	0.08	24.00	5.50	4.28	1.38	26.10	9.04
-6.98	33.83	4.73	7.06	18.60	10.90	25.54	0.11	39.80	9.22	3.00	1.13	44.81	6.65
-7.00	33.84	2.20	7.14	4.84	4.00	18.83	0.10	11.44	4.50	5.33	1.00	52.20	8.96
-6.99	33.84	4.20	7.45	16.80	11.98	25.20	0.07	32.10	16.70	2.95	0.30	87.00	6.64
-6.99	33.84	2.14	7.20	3.80	0.12	18.01	0.06	10.96	3.20	5.33	0.95	21.75	12.87
-7.01	33.84	4.05	7.33	12.52	15.14	28.80	0.07	28.40	20.33	5.55	0.25	20.88	7.75
-6.99	33.85	4.03	7.07	13.90	12.60	32.89	0.20	40.20	12.42	3.90	1.08	65.25	9.04
-7.00	33.85	4.64	7.04	10.00	6.60	26.76	0.07	24.10	12.33	3.98	1.03	13.05	9.29
-7.01	33.84	0.97	7.70	2.69	3.72	15.28	0.07	5.20	3.00	2.80	0.00	35.24	8.54
-7.03	33.85	2.07	7.12	3.00	4.50	20.63	0.14	11.10	2.89	6.25	2.03	0.00	10.65
-7.03	33.85	1.12	7.49	3.42	2.84	15.64	0.08	6.32	2.80	1.88	0.75	26.97	8.84
-7.02	33.85	1.33	7.18	2.84	3.40	20.14	0.22	7.12	5.90	2.45	0.88	3.48	11.40
-7.01	33.86	2.76	7.18	3.50	0.25	27.99	0.05	18.10	5.10	5.78	1.40	0.00	20.44
-7.01	33.87	7.39	7.14	13.60	12.90	60.69	0.17	53.20	24.80	5.28	1.85	1.74	16.67
-7.01	33.87	6.18	7.34	12.60	27.80	51.61	0.26	45.20	32.02	8.25	0.80	0.00	11.48
-7.01	33.86	2.11	8.24	4.60	2.96	22.32	0.09	10.60	4.47	8.05	0.85	0.00	11.48
-7.00	33.88	3.95	7.19	3.40	6.90	34.53	0.38	30.00	5.01	6.15	2.05	0.00	15.22
-7.01	33.87	2.90	6.20	1.00	3.90	29.21	0.08	20.60	11.00	1.35	0.00	11.31	18.66
-7.01	33.87	7.89	6.86	17.70	23.60	56.61	0.27	56.80	28.50	8.03	1.90	0.00	12.46
-7.02	33.87	3.73	7.48	6.00	3.10	34.53	0.19	21.40	13.96	4.63	1.88	3.48	16.19
-7.04	33.86	1.21	7.34	4.52	3.96	16.54	0.05	4.84	1.60	6.88	1.50	0.00	8.03
-7.05	33.86	2.72	7.40	4.95	0.45	22.88	0.04	15.30	5.50	4.53	1.65	7.40	13.92
-7.05	33.87	2.24	7.25	3.84	1.40	20.79	0.22	9.40	8.50	5.40	1.63	1.31	12.85
-7.06	33.87	2.31	7.37	1.90	1.10	20.83	0.15	12.30	4.20	4.78	1.70	1.74	17.01
-7.05	33.87	3.19	7.32	6.86	8.22	28.80	0.13	16.90	20.00	5.25	0.00	9.57	10.49
-7.04	33.86	3.28	7.56	8.00	5.10	27.17	0.05	15.60	15.60	6.75	0.73	256.65	10.62
-7.03	33.88	5.71	7.35	10.30	7.90	47.61	0.71	40.20	16.82	7.50	0.00	153.99	15.78
-7.02	33.88	2.65	7.70	4.30	1.70	24.72	0.08	15.00	10.32	3.33	1.15	143.55	14.27
-7.01	33.89	1.93	7.46	4.81	1.37	18.16	0.09	8.40	10.48	4.55	0.00	230.55	10.34
-7.07	33.87	4.89	7.80	8.30	5.20	39.44	0.21	32.50	12.50	5.70	0.83	62.21	15.18
-7.07	33.86	4.94	7.33	14.20	12.30	36.16	1.09	42.00	14.23	5.08	0.45	29.58	9.94
-7.10	33.85	5.60	7.54	15.40	5.60	41.07	0.29	42.60	13.51	3.23	1.03	95.70	12.67
-7.12	33.83	2.27	7.80	2.55	4.50	22.06	0.05	13.80	3.90	3.95	0.90	134.85	11.75
-7.05	33.85	1.70	7.60	1.95	3.36	17.26	0.09	7.10	1.90	7.40	0.78	6.09	10.59
-7.05	33.85	3.70	7.80	8.00	5.00	31.67	0.11	28.00	8.25	5.45	1.08	27.84	12.42
-7.10	33.83	3.11	7.70	3.50	0.10	31.26	0.32	21.40	6.40	4.70	0.90	42.20	23.30

-7.09	33.84	1.26	7.87	4.49	2.41	15.62	0.28	5.86	1.10	4.25	1.03	8.70	8.41
-7.07	33.83	1.83	8.08	1.40	0.10	18.99	0.05	10.00	3.10	5.38	1.08	7.40	21.93
-7.08	33.83	2.89	7.67	3.00	1.90	25.33	0.06	14.00	5.60	7.53	1.60	0.00	16.18
-7.04	33.84	3.13	7.55	6.75	1.15	30.03	0.06	19.00	9.30	7.00	0.90	6.09	15.11
-7.06	33.85	2.95	7.66	5.80	1.10	26.76	0.07	20.60	4.89	5.33	1.13	20.01	14.41
-7.04	33.86	2.37	7.45	4.80	1.70	22.06	0.11	9.00	8.00	9.73	0.88	45.24	12.24
-7.04	33.85	2.14	7.73	4.00	4.96	20.47	0.04	9.76	2.40	6.38	0.28	107.01	9.67
-7.03	33.85	0.90	7.62	1.32	1.46	13.68	0.06	2.72	0.60	3.98	0.78	47.85	11.60
-7.04	33.84	1.47	7.62	0.70	1.32	15.52	0.05	6.08	0.50	5.20	0.50	99.18	15.44
-7.05	33.83	6.21	7.74	12.50	8.50	50.06	0.36	47.00	15.50	6.10	0.45	19.14	15.45
-7.08	33.81	0.92	7.80	1.78	2.02	14.91	0.03	3.10	0.49	4.65	0.50	7.40	10.81
-7.09	33.81	1.27	7.65	0.48	1.00	15.64	0.04	4.36	0.00	6.83	0.73	35.67	18.18
-7.05	33.81	2.42	7.56	3.95	0.55	20.83	0.05	13.70	3.20	6.33	0.63	73.95	13.89
-7.04	33.81	1.35	7.57	0.55	0.45	14.66	0.04	5.02	0.00	6.35	0.88	46.11	20.73
-7.03	33.82	3.86	7.58	9.20	2.60	27.58	0.04	29.00	2.07	4.58	0.78	143.55	11.35
-7.01	33.82	2.38	7.57	7.30	0.20	17.56	0.07	15.30	3.50	4.00	0.83	158.34	9.07
-7.04	33.78	1.99	7.48	4.40	2.70	18.38	0.16	11.80	1.56	5.10	0.75	96.57	9.76
-7.04	33.78	1.70	7.40	2.47	0.56	17.15	0.05	7.96	0.50	5.70	0.63	27.84	13.94
-7.03	33.80	1.30	7.50	0.50	0.55	14.33	0.01	5.30	0.00	4.55	0.88	100.92	19.78
-7.02	33.81	5.30	7.26	17.90	14.10	33.71	0.08	43.40	16.24	3.75	0.40	62.21	8.43
-7.02	33.81	8.08	7.66	3.70	7.30	76.43	0.44	63.40	15.80	5.25	0.88	76.56	32.59
-7.02	33.81	1.68	7.45	2.35	1.53	17.36	0.16	9.04	1.50	4.85	0.65	47.85	12.46
-7.02	33.81	1.54	7.56	3.61	1.69	20.02	0.23	8.50	1.90	3.78	0.58	43.50	12.30
-7.01	33.81	4.34	7.33	11.90	9.10	27.99	0.14	34.00	8.31	4.08	0.75	70.47	8.64
-7.00	33.82	1.94	7.46	3.95	0.55	16.95	0.03	12.20	2.40	5.13	0.75	36.11	11.30
-6.98	33.82	5.23	7.34	20.10	7.90	31.26	0.19	40.20	12.70	3.58	0.98	97.88	8.35
-6.98	33.81	4.16	7.31	17.84	6.28	27.20	0.05	31.30	15.38	2.70	0.00	116.58	7.83



## Master ST CAC Agiq

Mémoire de fin d'études pour l'obtention du Diplôme de Master Sciences et Techniques

Nom et prénom: **HAMID Rajaa**

Année Universitaire : 2017/2018

**Titre: Application de la régression logistique à l'évaluation de la qualité chimique des eaux d'irrigation dans la région de Skhirat**

### *Résumé*

Notre travail a pour but d'évaluer les paramètres physico-chimiques de l'eau d'irrigation dans la région de Skhirat qui est caractérisée par une activité maraîchère intense. Cette évaluation est faite à partir des méthodes d'analyse statistique; la régression logistique.

Des échantillons d'eau sont prélevés à partir de 70 puits d'une façon représentative de la région. Ensuite, d'analyses des paramètres physico-chimiques de ces échantillons sont réalisées. Les résultats obtenus sont organisés dans un tableau appelé matrice des données.

Le traitement des données par la régression logistique montre que les eaux souterraines dans la région présentent un risque moyen de salinité avec un faible risque d'alcalinité.

L'analyse des données par nos modèles montre que les modèles de prédiction de la CE par la régression logistique binaire et ordinaire et SAR par la régression logistique binaire ont une très bonne qualité Prédictive.

Alors que le modèle de régression logistique ordinaire pour le SAR a une qualité prédictive moyenne.

**Mots clés:** Paramètres, physico-chimique, l'eau d'irrigation, Skhirat, Régression logistique.